

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Patrick M. ALLIEL, et al.

SERIAL NO.: NEW U.S. PCT APPLICATION

FILED: HEREWITH

INTERNATIONAL APPLICATION NO.: PCT/FR99/01513

INTERNATIONAL FILING DATE: 23 June 1999

FOR: NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY WITH
HUMAN ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES**REQUEST FOR PRIORITY UNDER 35 U.S.C. 119**
AND THE INTERNATIONAL CONVENTIONAssistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In the matter of the above-identified application for patent, notice is hereby given that the applicant claims as priority:

<u>COUNTRY</u>	<u>APPLICATION NO.</u>	<u>DAY/MONTH/YEAR</u>
FRANCE	98/07920	23 June 1998

A certified copy of the corresponding Convention application(s) was submitted to the International Bureau in PCT Application No. **PCT/FR99/01513**. Receipt of the certified copy(s) by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.

Respectfully submitted,
OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.



Norman F. Oblon
Attorney of Record
Registration No. 24,618
Surinder Sachar
Registration No. 34,423

**22850**

RECEIVED

1971 JUN 15 10 10 AM - 047850



PCT/FR 39 / 01513

23 JUIN 1999

REC'D 18 AUG 1999

WIPO PCT

BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE 09/719554

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 29 JUIN 1999

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

SIEGE
26 bis, rue de Saint Petersburg
75800 PARIS Cédex 08
Téléphone : 01 53 04 53 04
Télécopie : 01 42 93 59 30

DOCUMENT DE
PRIORITE
PRESENTE OU TRANSMIS
CONFORMEMENT A LA REGLE
17.1.a) OU b)



REQUÊTE EN DÉLIVRANCE

Confirmation d'un dépôt par télécopie ☐

Cet imprimé est à remplir à l'encre noire en lettres capitales

Réservé à l'INPI

DATE DE REMISE DES PIÈCES

N° D'ENREGISTREMENT NATIONAL

23 JUN 1998

DÉPARTEMENT DE DÉPÔT

98 07920 -

DATE DE DÉPÔT

23 JUIN 1998

2 DEMANDE Nature du titre de propriété industrielle

☒ brevet d'invention

☐ demande divisionnaire

☐ certificat d'utilité

☐ transformation d'une demande de brevet européen

☐ demande initiale

☐ brevet d'invention

n° du pouvoir permanent

références du correspondant
BLOcp598/21FR

téléphone

date

Établissement du rapport de recherche

☐ différé

☒ immédiat

Le demandeur, personne physique, requiert le paiement échelonné de la redevance

☐ oui

☐ non

Titre de l'invention (200 caractères maximum)

FAMILLE DE SEQUENCES NUCLEIQUES ET DE SEQUENCES PROTEIQUES DEDUITES
PRESENTANT DES MOTIFS RETROVIRAUX ENDOGENES HUMAINS ET LEURS
APPLICATIONS.

3 DEMANDEUR (S)

n° SIREN

code APE-NAF

Nom et prénoms (souligner le nom patronymique) ou dénomination

INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE
MEDICALE-INERM

Forme juridique

Etablissement public

Nationalité (s) française

Adresse (s) complète (s)

101 rue de Tolbiac, 75654 PARIS Cedex 13

Pays

FRANCE

En cas d'insuffisance de place, poursuivre sur papier libre ☐

4 INVENTEUR (S) Les inventeurs sont les demandeurs

☐ oui

☒ non

Si la réponse est non, fournir une désignation séparée

5 RÉDUCTION DU TAUX DES REDEVANCES

☐ requise pour la 1ère fois

☐ requise antérieurement au dépôt ; joindre copie de la décision d'admission

6 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE

pays d'origine

numéro

date de dépôt

nature de la demande

7 DIVISIONS

antérieures à la présente demande

n°

date

n°

date

8 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE

(nom et qualité du signataire - n° d'inscription)

Béatrice ORES
n° 92-4046

SIGNATURE DU PRÉPOSÉ À LA RÉCEPTION

SIGNATURE APRÈS ENREGISTREMENT DE LA DEMANDE À L'INPI

DÉSIGNATION DE L'INVENTEUR

(si le demandeur n'est pas l'inventeur ou l'unique inventeur)

DIVISION ADMINISTRATIVE DES BREVETS

26bis, rue de Saint-Petersbourg
75800 Paris Cédex 08
Tél. : 01 53 04 53 04 - Télécopie : 01 42 93 59 30

N° D'ENREGISTREMENT NATIONAL

98 07920

TITRE DE L'INVENTION :

FAMILLE DE SEQUENCES NUCLEIQUES ET DE SEQUENCES PROTEIQUES DEDUITES
PRESENTANT DES MOTIFS RETROVIRAUX ENDOGENES HUMAINS ET LEURS APPLICATIONS.

LE(S) SOUSSIGNÉ(S)

CABINET ORES
6, Avenue de Messine
75008 PARIS

DÉSIGNE(NT) EN TANT QU'INVENTEUR(S) (indiquer nom, prénoms, adresse et souligner le nom patronymique) :

ALLIEL Patrick M.

4, rue Lazare Carnot
92140 CLAMART (FRANCE)

PERIN Jean-Pierre

182, rue d'Aulnay
92350 LE PLESSIS-ROBINSON (FRANCE)


RIEGER François

38bis Boulevard de la République
92100 BOULOGNE (FRANCE)

NOTA : A titre exceptionnel, le nom de l'inventeur peut être suivi de celui de la société à laquelle il appartient (société d'appartenance) lorsque celle-ci est différente de la société déposante ou titulaire.

Date et signature (e) du (des) demandeur (e) ou du mandataire

Paris, le 21 Décembre 1998


Béatrice ORES
Mandataire
N° 92-4046

DOCUMENT COMPORTANT DES MODIFICATIONS

PAGE(S) DE LA DESCRIPTION OU DES REVENDEICATIONS OU PLANCHE(S) DE DESSIN			R.M.*	DATE DE LA CORRESPONDANCE	TAMPON DATEUR DU CORRECTEUR
Modifiée(s)	Supprimée(s)	Ajoutée(s)			
6'				R. 11.98	16 NOV. 1998 - SA
73 et 76			X	15.01.99	19 JAN. 1999 - SR

La présente invention est relative à une nouvelle famille de séquences nucléiques et de séquences protéiques déduites, qui présentent des motifs rétroviraux endogènes humains complets ou partiels.

L'invention est également relative à la détection et/ou à l'utilisation
5 desdites séquences nucléiques et desdites séquences protéiques correspondantes, dans le cadre d'applications diagnostiques, prophylactiques et thérapeutiques, en particulier pour des neuropathologies à composante autoimmune comme la sclérose en plaques.

L'invention concerne aussi l'obtention de sondes nucléiques double
brins et simple brin anti-sens, de ribozymes, aptes à moduler la réplication virale (T.R.
10 Cech, Science, 1987, 236, 1532-1539 ; R.H. Symons, *Trends Biochem. Sci.*, 1989, 14, 445-450) des molécules recombinantes correspondantes, et des anticorps associés.

Les rétrovirus sont des virus qui se répliquent uniquement en utilisant la voie inverse du traitement classique de l'information génétique. Ce processus, nommé transcription inverse, est médié par une ADN polymérase ARN dépendante ou
15 transcriptase reverse, codée par le gène *pol*. Les rétrovirus codent aussi au minimum pour deux gènes additionnels. Le gène *gag* code pour les protéines du squelette, matrice, nucléocapside et capsid. Le gène *env* code pour les glycoprotéines d'enveloppe. La transcription rétrovirale est régulée par des régions promotrices ou "enhancers", situées dans des régions hautement répétées ou LTR (*Long Terminal Repeat*) et qui
20 sont présentes aux deux extrémités du génome rétroviral.

Lors de l'infection d'une cellule, la polymérase fait une copie ADN du génome ARN ; cette copie peut alors s'intégrer dans le génome humain. Les rétrovirus ne tuent pas les cellules qu'ils infectent, mais au contraire améliorent souvent leur rapidité de croissance. Les rétrovirus peuvent infecter des cellules germinales ou
25 des embryons à un stade précoce ; ils peuvent dans ces conditions, intégrer la lignée germinale et être transmis par transmission mendélienne verticale, ce qui constitue la relation la plus étroite entre un hôte et son parasite. Ces virus endogènes peuvent dégénérer au cours des générations de l'organisme hôte et perdre leurs propriétés initiales. Cependant certains d'entre eux peuvent conserver tout ou partie de leurs propriétés
30 ou des propriétés des motifs les composant, ou encore acquérir de nouvelles propriétés fonctionnelles présentant un avantage pour l'organisme hôte, ce qui expli-

querait la préservation de leur séquence.

L'existence de motifs endogènes présentant de longs cadres de lecture ouverts et/ou soumis à une forte pression de sélection peut donc être indicatrice d'une fonction biologique préservée ou acquise, qui peut correspondre à un bénéfice pour l'organisme hôte. Ces séquences rétrovirales peuvent aussi subir, au cours des générations, des modifications discrètes qui vont être à même de réveiller certaines de leurs potentialités et engendrer ou favoriser des processus pathologiques. Il est apparu récemment nécessaire de faire le bilan et d'identifier ces séquences afin de pouvoir évaluer leur impact fonctionnel.

Les séquences rétrovirales endogènes humaines ou HERVs représentent une part importante du génome humain. Ces régions rétrovirales se présentent sous plusieurs formes :

- des structures rétrovirales endogènes complètes associant des motifs *gag*, *pol* et *env*, flanqués de séquences nucléiques répétées, qui montrent une analogie significative avec la structure LTR-*gag-pol-env*-LTR des rétrovirus infectieux,

- des séquences rétrovirales tronquées ; par exemple, les rétrotransposons sont privés de leur domaine *env* et les rétroposons ne possèdent pas les régions *env* et LTR.

Jusqu'à présent l'étude de ces régions du génome a été négligée chez l'Homme pour deux raisons essentielles :

- l'existence d'insertions/délétions qui peuvent décaler le cadre de lecture et de mutations qui modifient la séquence. Ces modifications entraînent des altérations de la structure et par conséquent de la fonction biologique de ces motifs.

- l'absence d'associations avérées avec des pathologies humaines.

La connaissance, récente de fragments significativement représentatifs du génome humain et une orientation des recherches vers une étude structure/fonction des motifs rétroviraux endogènes, ont permis de préciser l'intérêt de ces régions. L'implication de séquences endogènes tronquées ou complètes dans des pathologies chez l'animal est documentée ; par exemple leur association avec des processus tumoraux a été clairement mise en évidence (S.K. Chattopadhyay et coll.,

1982, *Nature*, **295**, 25-31). Une recherche visant à préciser l'association ou l'influence des HERVs dans des pathologies humaines se justifie donc aujourd'hui.

Une classification des éléments HERV a été proposée (Tönjes R.R. et al., *AIDS & Hum. Retrovirol.*, 1996, **13**, S261-S267; A.M. Krieg et al., *FASEB J.*, 1992, **6**, 2537-2544). Elle est basée sur une homologie de ces séquences avec des rétrovirus isolés chez les animaux, à l'aide de sondes rétrovirales hétérologues. En effet, en général, les HERVs présentent relativement peu d'homologie avec des rétrovirus infectieux humains connus.

Les familles de classe I présentent une homologie de séquence avec les rétrovirus de mammifères de type C ; on peut citer notamment la superfamille ERI, proche du virus MuLV (*murine leukemia virus*) et du virus BaEV (*baboon endogenous virus*).

Les familles de classe II présentent une homologie de séquence avec les rétrovirus de mammifères de type B tel que le MMTV (*mouse mammary tumour virus*) ou les rétrovirus de type D tel que le SRV (*squirrel monkey retrovirus*).

D'autres familles ont également été décrites ; parmi celles-ci, on peut citer des HERVs qui présentent, de manière exceptionnelle, une homologie partielle avec HTLV-1 (RTVL-H) ou des virus de primates ; HRES-1, par exemple, présente une homologie de séquence avec des HTLVs.

Les programmes de très grand séquençage du génome humain permettent aujourd'hui de disposer d'un nombre significatif de nouvelles séquences rétrovirales. L'usage de logiciels de traitement de données permet d'identifier et d'analyser ces gènes. Dans ce contexte une recherche systématique portant sur l'ensemble des informations disponibles à ce jour a été engagée afin d'identifier de nouvelles séquences rétrovirales endogènes humaines en fonction de certains critères d'analyse :

- présence de longs cadres de lecture ouverts conservés au cours de l'évolution de l'organisme hôte et pouvant laisser envisager une fonction biologique,
- analogie avec des séquences déjà caractérisées en dehors ou dans le domaine des rétrovirus,

- localisation dans des régions de susceptibilité pour certaines pathologies ou à proximité de gènes essentiels, par exemple dans les domaine du cancer, des

régulation du système immunitaire ou dans certaines neuropathologies.

Les recherches effectuées par les Inventeurs, dans des bases de données de séquences leur ont permis d'identifier un ensemble de séquences ou de motifs rétroviraux endogènes dont l'expression normale ou pathologique peut favoriser ou perturber un effet protecteur vis-à-vis de processus pathologiques, ou intervenir dans le déclenchement ou l'aggravation de pathologies.

La présente invention a pour objet un fragment d'acide nucléique purifié, caractérisé en ce qu'il comprend tout ou partie d'une séquence codant pour une séquence rétrovirale endogène humaine, qui présente au moins des motifs rétroviraux de type *env*, répondant à la séquence SEQ ID NO:1 ou à une séquence présentant un niveau d'homologie avec ladite séquence SEQ ID NO:1 supérieur ou égal à 80% sur plus de 190 nucléotides ou supérieur ou égal à 70 % sur plus de 600 nucléotides pour les domaines de type *env*.

On entend par séquence homologue, aussi bien une séquence qui présente une identité complète ou partielle avec la séquence SEQ ID NO:1 précitée qu'une séquence qui présente une similarité partielle avec ladite séquence SEQ ID NO:1.

Selon un mode de réalisation avantageux dudit fragment, il présente à la fois des motifs rétroviraux correspondant à un domaine *env* et répondant à la séquence SEQ ID NO:1 et des motifs rétroviraux correspondant à un domaine *gag* et répondant à la séquence SEQ ID NO:2 ou à une séquence présentant un niveau d'homologie supérieur ou égal à 80 % sur plus de 190 nucléotides ou supérieur ou égal à 70 % sur plus de 600 nucléotides pour les domaines de type *env* et un niveau d'homologie supérieur ou égal à 90 % sur plus de 700 nucléotides ou supérieur ou égal à 70 % sur plus de 1200 nucléotides pour les domaines de type *gag*, lesquels motifs ne présentent aucune insertion ou délétion de plus de 200 nucléotides.

Lesdits fragments constituent une nouvelle famille de séquences rétrovirales endogènes humaines (famille HERV-7q) qui présente une homologie de séquence avec les rétrovirus MSRV, tels que décrits dans la Demande Internationale WO 97/06260 ; lesdits fragments selon la présente invention présentent :

- deux motifs nucléotidiques répétés de 711 pb (figure 3), présentant des signaux caractéristiques identifiés dans des LTRs (*Long Terminal Repeats*): promoteurs de transcription de type boîtes TATAA ou CCAAT. Ces domaines répétés encadrent trois motifs déduits de type-*gag*, *pol* et *env* (figure 2).

5 - un motif de type *env* (positions 6965 nt - 9550 nt sur la séquence SEQ ID NO:3) qui contient un long cadre de lecture ouvert de 1620 nucléotides (positions 7874-9493 de la séquence ID NO:3), codant pour une protéine de séquence inédite de 540 acides aminés (figure 4) et fragment souligné de la SEQ ID NO:27. On retrouve à l'intérieur du domaine trans-membranaire de ce domaine *env*, un motif
10 peptidique de type CKS-25/CKS-17 (fig.5), reconnu pour présenter des fonctions immunosuppressives sur les cellules lymphocytaires hôtes (M. Mitani et coll., 1987, *Proc. Natl. Acad. Sci. USA*, 84, 237-240). Un domaine de type zinc-finger HX_3 - $HX_{22-33}CX_2C$ (Kulkolski et coll., 1992, *Mol. Cell. Biol.*, 12, 2331-2338), que l'on
15 retrouve dans des domaines de type intégrase est identifié dans un autre cadre de lecture. Ce domaine *env* particulier signe la caractéristique de nouveaux motifs rétro-viraux endogènes.

 - le motif (positions 3065 nt - 4390 nt sur la séquence SEQ ID NO:3) de type-*gag* codant pour des motifs protéiques selon la figure 6 (SEQ ID NO:51) (positions 3118-4198 de la SEQ ID NO:3) a été identifié grâce à des analogies avec
20 des domaines *gag* connus. On retrouve, par exemple, la région d'homologie majeure QX_3EX_7R (Benit et coll., 1997, *J. Virol.*, 71, 5652-5657). Le motif de fixation des acides nucléiques $CX_2CX_3_4HX_4C$, situé en position C-terminale, est identifié dans un autre cadre de lecture (Covey et coll., 1986, *Nucleic Acids Res.*, 14, 623-633). En
25 amont du domaine *gag* on détecte un motif de 182 nucléotides répété deux fois (figure 1).

 - le domaine *pol* présente les consensus classiques d'une région *pol* de rétrovirus au niveau des domaines protéase, transcriptase reverse et RNase H. On retrouve dans *pol* un motif proche du consensus LLDTGA (Weber et coll., 1988, *Science*, 243, 928-931). Les motifs D et AF, LPQ et SP, et YVDD (Xiong et
30 Eickbush, 1990, *EMBO J.*, 9, 3353-3362), sont respectivement retrouvés dans les 3°, 4° et 5° boîtes d'homologie. Les motifs YTDGSS et TDS sont présents dans la région

de la RNase H,

- les régions *gag* et *pol* pourraient être considérées comme jointives avec un passage de la région *gag* à la région *pol* par un décalage du cadre de lecture.

La présente invention englobe les séquences appartenant à la famille
 5 HERV-7q telle que définie ci-dessus (présence de la séquence SEQ ID NO:1 ou d'une séquence homologue ou présence à la fois des séquences SEQ ID NO:1 et SEQ ID NO:2) et notamment les séquences SEQ ID NO:3-21; elle englobe également les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes ainsi que les fragments issus des régions codantes des
 10 séquences précédentes correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires. (SEQ ID NO :30-50)

Ces différents fragments peuvent avantageusement être utilisés comme amorces ou comme sondes; ils s'hybrident spécifiquement à une séquence de la famille HERV-7q.

15 Parmi ces fragments, on peut citer, de préférence les fragments suivants:

- un fragment de 182 nucléotides répété deux fois, situé en amont du domaine *gag* aux positions 2502-2611/2613-2865 de la SEQ ID NO:3;

Amorces et sondes spécifiques de la région *gag*

20 - une amorce G1F, sens, localisée dans la région amont du domaine *gag* de HERV-7q : 5' GGACCATAGAGGACACTCCAGGACTA 3' (SEQ ID NO:30);

- une amorce G1R, anti-sens, localisée dans la région 3' terminale du domaine *gag* : 5' CCTCAGTCCTGCTGCTGGATCATCT 3' (SEQ ID NO :31)

25 - le fragment de 1505 nt amplifié par le couple G1F-G1R est utilisé afin de générer les sondes aptes à hybrider les différents produits d'amplification des PCR;

- une amorce G2F, sens nichée : (SEQ ID NO :32)

5' CCTCCAAGCAGTGGGAGGAAGAGAATT 3'

- une amorce G2R, anti-sens nichée : (SEQ ID NO :33)

5' CCTTCCCTGTGTTATTGTGGACATCATT 3'

- une amorce G4F, sens nichée : (SEQ ID NO :34)

5' GGAAGAAGTCTATGAATTATTCAATGATGT 3'

5

- une amorce G3F, sens nichée : (SEQ ID NO :35)

5' GGGACACAGAATCAGAACATGGAGATT 3'

- une amorce G4R, anti-sens nichée : (SEQ ID NO :36)

5' GCCTTCAGAAGAGTCAGGTGACAGAGA 3'

- une amorce G5R, anti-sens nichée : (SEQ ID NO :37)

10

5' GAGCCTCCAAAGTCCACTTGCCTGA 3'

Amorces et sondes spécifiques de la région env

- une amorce E1F, sens : (SEQ ID NO :38)

5' GATTTTCAGTATCTACTAGTCTGGGTAGAT 3'

- une amorce E1R, anti-sens : (SEQ ID NO :39)

15

5' CTAGGAAATCCAGCTAGTCCTGTCTCA 3'

- le fragment de 2529 nt amplifié par le couple d'amorces E1F-E1R, est utilisé afin de générer les sondes aptes à hybrider les différents produits d'amplification des PCR.

- une amorce E2F, sens : (SEQ ID NO :40)

20

5' CCAAGACAGCCAAGTTAGTTGCAGACAT 3'

- une amorce E2R, antisens : (SEQ ID NO :41)

5' GGACGCTGCATTCTCCATAGAACTCTT 3'

- une amorce E3F, sens : (SEQ ID NO :42)

5' GCAATACTACATACACAACCAACTCCCAA 3'

25

- une amorce E3R, anti-sens : (SEQ ID NO :43)

5' GGGGGAGGCATATCCAACAGTTAGTA 3'

- une amorce E4F, sens : (SEQ ID NO :44)

5' CCATCTACACTGAACAAGATTTATACACTT 3'

- une amorce E4R, anti-sens : (SEQ ID NO :45)

30

5' AATGCCAGTACCTAGTGCACCTAGCACT 3'

- une amorce E5F, sens : (SEQ ID NO :46)

5' CGAATACAACGTAGAGCAGAGGAGCTTCGAA 3'

- une amorce E6F, sens : (SEQ ID NO :47)

5' AGCCCAAGATGCAGTCCAAGACTAAGAT 3'

5

- une amorce E5R : (SEQ ID NO :48)

5' GCGTAGTAGAGGTTGTGCAGCTGAGAT 3'

- une amorce ExF : (SEQ ID NO:49)

CCCTTACCAAGAGTTTCTATGGAGAAT

- une amorce ExR : (SEQ ID NO:50)

10

ACCGCTCTAACTGCTTCCTGCTGAATT

Tous les oligonucléotides sont conçus pour pouvoir générer une amorce sens et une amorce anti-sens par un décalage de la séquence de l'amorce de référence de 1 à 7 nucléotides vers le côté 5' ou vers le côté 3': la modification de la séquence peut entraîner une modification de la taille de l'amorce de 1 à 7 nucléotides selon les cas. Les amorces choisies peuvent être optimisées selon les cas par un raccourcissement ou un allongement portant sur 1 à 9 nucléotides.

De manière préférée, l'hybridation, le clonage, le sous-clonage, l'obtention, la préparation et l'analyse des acides nucléiques, des peptides et des anticorps, le séquençage des acides nucléiques et des peptides, l'hybridation *in situ* et l'immunohistochimie sont réalisés dans les conditions décrites dans les ouvrages suivants :

- Current Protocols in Molecular Biology. Eds. F.M Ausubel, R. Brent & R.E Kingston et coll. Green Publishing associates and Wiley Interscience.
- Molecular Cloning: a laboratory manual. Eds. J. Sambrook, E.F. Fritsch & T. Maniatis. Cold Spring Harbor Laboratory Press. Cold Spring Harbor.
- The Practical Approach series. Eds. D. Rickwood & B.D. Ames. IRL Press and Oxford University Press. En particulier, antibodies I & II; DNA cloning I, II, III; Nucleic acid and protein sequence analysis; Nucleic acid hybridization; Nucleic acid sequencing ; Oligonucleotide synthesis; Protein purification applications; Protein purification methods; Protein sequencing; Transcription and translation; Gels electrophoresis of nucleic acids; Gels electrophoresis of proteins; Genome analysis;

HPLC of macromolecules; Human genetic diseases; Microcomputing in biology; Molecular neurobiology; Mutagenicity testing; Essential molecular biology I & II.

- Proteome research: New frontiers in functional genomics. Eds M.R. Wilkins & coll.. Springer.

5 La séquence rétrovirale endogène humaine (SEQ ID NO:3), située sur le bras long du chromosome 7 correspond à la séquence HERV-7q ; elle présente 10,5 kb (fig. 1 et 2) et répond aux critères précédemment définis.

La recherche de domaines présentant des similitudes, tout ou partie, avec les régions *gag* et *env* de HERV-7q a abouti à l'identification de nouvelles
10 séquences rétrovirales endogènes. Ces séquences peuvent présenter la structure d'un rétrovirus endogène complet comme la séquence rétrovirale endogène située à proximité du gène des sous-unités alpha et delta du récepteur des cellules-T, et dénommée en conséquence HERV-TcR ; à titre d'exemple la figure 7 montre la comparaison des alignements nucléiques des domaines *gag* respectifs de HERV-7q et HERV-TcR
15 (séquence HG12, SEQ ID NO:18). On trouve aussi des structures rétrovirales partielles. Ces domaines rétroviraux similaires à HERV-7q sont identifiées dans des séquences nucléiques indépendantes comme le montre leur localisation chromosomique. Des motifs nucléiques (appelés ici, HEx ou HGx et respectivement analogues à des domaines de type *env* ou *gag*) ressemblant aux domaines *env* ou *gag* de
20 HERV-7q ont été retrouvés, à l'aide des banques de données précitées :

- HE2 : chromosome 17 (SEQ ID NO:4),
- HE3 et HG3: chromosome 6 (SEQ ID NO:5 et 6),
- HE4 : chromosome X (SEQ ID NO:7),
- HE5 : chromosome X q22 (SEQ ID NO:8),
- 25 - HE6 et HG6 : chromosome 1 q23.3-q24.3 (SEQ ID NO:9 et 10),
- HE7 : chromosome 7 p15 (SEQ ID NO:11),
- HE8 et HG8 : chromosome 19 (SEQ ID NO:12 et 13),
- HE9 : chromosome X (SEQ ID NO:14),
- HE 10 : chromosome X q13.1-21.1 (SEQ ID NO:15),
- 30 - HE11 et HG11 : chromosome 7 q21-22 (SEQ ID NO:16 et 17),

- HE12 et HG12, dans HERV-TcR : chromosome 14 q11.2 (SEQ ID NO:18 et 19).

Les alignements des domaines *env* (fig. 8) et *gag* (fig. 9) explicitent les niveaux d'homologie observés entre les séquences décrites ci-dessus et les séquences homologues dans HERV-7q. Les analogies peuvent s'étendre aux motifs rétroviraux flanquants.

Une analyse des séquences étiquettes disponibles dans les banques de données montre que des transcrits appartenant à certains des membres de cette famille, en particulier HERV-7q, s'expriment essentiellement dans des tissus d'origine foétale ou placentaire.

Des séquences polypeptidiques générées par ces transcrits peuvent donc être potentiellement produites et des fonctions ou activités biologiques peuvent être envisagées, par analogie avec des polypeptides biologiquement actifs d'origine virale ou rétrovirale ; par exemple, les motifs peptidiques de type CKS-17 (fig. 5) ou CKS-25 (Huang S.S et Huang J.S, J. Biol. Chem. 1998, 273, 4815-4818), qui présentent des fonctions immunomodulatrices sur les cellules lymphocytaires hôtes. Les différences de séquence observées et d'éventuelles modifications normales ou pathologiques, sont en particulier, à l'origine d'une modulation de la fonction.

HERV-7q représente le paradigme de la nouvelle famille de séquences rétrovirales endogènes humaines ou de motifs rétroviraux endogènes.

HERV-7q et certaines des séquences rétrovirales endogènes appartenant à sa famille, présentent un domaine de type *pol* analogue à des séquences rétrovirales de type *pol* comme par exemple la région *pol* identifiée dans le rétrovirus MSRV associé à la sclérose en plaques et décrit par H. Perron et al. (1997, *Proc. Natl. Acad. Sci. USA*, 94, 7583-7588 ; Demande Internationale PCT WO 97/06260).

Toutefois, les séquences selon la présente invention se distinguent des séquences rétrovirales exogènes infectieuses analogues à MSRV antérieurement décrites en ce que les séquences *gag* et *env*, selon l'invention sont significativement différentes selon les critères précédemment définis et en fonction de certaines caractéristiques spécifiques, par exemple le long cadre de lecture ouvert du domaine *env* de HERV-7q ; elles seraient à même de permettre de signer une pathologie lorsqu'elles

présentent des insertions, des délétions, des décalages de cadre de lecture ou des mutations.

En effet, les différences observées entre les séquences humaines de type HERV-7q, qui sont isolées d'individus réputés normaux et les séquences issues de certains échantillons d'origine pathologique, ne sont pas distribuées au hasard. Des comparaisons menées entre la région *gag* provenant de particules rétrovirales infectieuses (N° d'accèsion EMBL: A60168, A60200, A60201, A60171...) et la séquence *gag* correspondante de HERV-7q (fig. 9), permettent d'observer que les mutations affectent préférentiellement des codons non-sens. Par exemple, deux codons non-sens dans HERV-7q sont remplacés par un codon arginine dans A60200, ce qui permet d'obtenir une séquence déduite de 109 acides aminés pour HERV-7q et de 166 acides aminés pour A60200. Les changements de base permettent en conséquence de prolonger le cadre de lecture et de coder potentiellement pour des structures polypeptidiques de plus grande taille (figure 10).

De même, une séquence de type *env* provenant de particules rétrovirales infectieuses, présente une analogie significative avec le domaine *env* de HERV-7q (figure 11). Ces analogies marquées entre séquences rétrovirales exogènes et endogènes pourraient être à l'origine du déclenchement ou de l'aggravation de certains processus pathologiques, en particulier de certaines maladies auto-immunes, comme la sclérose en plaques. A cet égard, on peut remarquer que certaines des séquences rétrovirales endogènes décrites dans l'invention se situent à proximité ou dans des régions réputées présenter une susceptibilité pour la sclérose en plaques : par exemple HERV-7q et la région 7q21-22 du chromosome 7, de même pour HE12 et HG12 dans HERV-TcR et la région du gène codant pour les chaînes alpha et delta du récepteur des cellules-T, HE2 et le chromosome 17, ou HE3 et HG3 et le chromosome 6.

On n'observe aucune homologie significative avec des séquences rétrovirales endogènes déjà décrites; par contre, on peut relever une homologie limitée et en tout état de cause inférieure aux critères définis selon l'invention entre les domaines *env* de la séquence HERV-7q (SEQ ID NO :1) et de la séquence HERV-9 (figure 12). La figure 13 montre des homologies étendues entre la séquence HERV-7q

avec une séquence rétrovirale exogène (N° d'accension EMBL : A60170).

Les séquences rétrovirales endogènes humaines appartenant à la famille de HERV-7q, peuvent protéger contre des agressions liées à l'environnement ou constituer un bénéfice pour l'individu. Cet effet bénéfique pourrait être une des
 5 raisons possibles de la pression de sélection exercée sur certaines de ces séquences et du caractère potentiellement fonctionnel des structures protéiques déduites identifiées : par exemple le long cadre de lecture ouvert apte à coder pour une nouvelle protéine et correspondant au domaine *env* de HERV-7q.

Les séquences rétrovirales endogènes humaines appartenant à la famille de HERV-7q pourraient être associées par exemple, à des pathologies en relation
 10 avec les processus liés au cancer, aux neuropathologies à composante auto-immune ou à tout autre processus pathologique en association ou non avec des virus ou rétrovirus endogènes ou exogènes. Leur action pourrait porter sur la déclaration, l'aggravation, la modification du calendrier d'apparition ou encore la protection vis à vis de la maladie.

Dans le contexte d'application à des pathologies autoimmunes
 15 (comme par exemple le lupus, le syndrome de Sjögren, la polyarthrite rhumatoïde, la sclérose en plaques...), on peut relever des analogies significatives entre les motifs rétroviraux endogènes identifiés et des motifs retrouvés dans des structures rétrovirales caractérisées chez des patients présentant des pathologies autoimmunes comme la
 20 sclérose en plaques : par exemple des fragments de domaine *gag* (récemment disponibles dans les banques de données) provenant de particules rétrovirales infectieuses ou encore la séquence complète du domaine *pol* correspondant au virus MSRV associé à la sclérose en plaques. Ces motifs rétroviraux possèdent des analogies significatives avec les séquences endogènes homologues de type HERV-7q, ce qui permet
 25 d'envisager une association directe ou indirecte avec des processus pathologiques, dont la sclérose en plaques, en association ou non avec MSRV. On peut relever la présence de certaines de ces séquences ou motifs dans des régions de susceptibilité pour la sclérose en plaques: par exemple, les séquences HE11 et HG11, autour de la région 7q 21-22 ou encore HE4, HE5, HE6, HE9, HE10 ou HG10 sur le chromosome
 30 X sont localisées au niveau ou à proximité de régions chromosomiques régulièrement associées à des gènes de susceptibilité pour la sclérose en plaques. Ces séquences

seraient donc à même de fournir des moyens de localisation ou d'identification des gènes de prédisposition.

L'intérêt de ces séquences dépasse le cadre des maladies auto-immunes. En dehors de l'importance générale des motifs rétroviraux dans le déclenchement ou l'aggravation d'un processus tumoral, bien montré en particulier dans les modèles murins (H. Fan dans *The retroviridae*, 1994, ed. J.A. Levy, Plenum, New York, p. 313-353), ces séquences pourraient se retrouver à proximité ou au sein de gènes importants et en altérer l'expression : par exemple HERV-TcR et les gènes des sous-unités alpha et delta du récepteur des cellules-T impliquées dans des perturbations de la fonction immunitaire.

L'invention a également pour objet les transcrits générés à partir des séquences précitées ainsi que celles présentant éventuellement des modifications avec les séquences de référence décrites dans l'invention lorsqu'ils sont exprimés chez certains patients.

En effet, les systèmes de régulation de l'expression des protéines rétrovirales de HERV-7q, qui sont présents dans les motifs de type LTR, pourraient influencer l'expression de gènes situés dans le voisinage chromosomique proche ou éloigné et induire des perturbations à caractère immunologique et/ou neurologique. Par exemple la séquence rétrovirale endogène HERV-TcR, se trouve à proximité immédiate des gènes des sous-unités alpha et delta du récepteur des cellules-T précédemment décrit. Les motifs de type LTR pourraient aussi coder pour des superantigènes (Acha-Orbea et Palmer, 1991, *Immunol. Today*, 12, 356-361). D'une manière générale des protéines rétrovirales de type HERV-7q ou apparenté, ou leurs formes tronquées ou partielles pourraient être impliquées dans des phénomènes de cytotoxicité ou de superantégenité, comme par exemple celles issues du long cadre de lecture ouvert identifié dans le domaine *env* (figure 4).

A cet égard, on peut relever que des motifs rétroviraux issus de régions déficientes sont aptes à présenter des fonctions biologiques: par exemple, la protéine d'enveloppe p15E issue de motifs rétroviraux déficients, possède une activité anti-inflammatoire et immunosuppressive (Snyderman et Ciancolo, 1984, *Immunol. Today*, 5, 240-244).

Ces structures sont vraisemblablement à même de provoquer des brèches ou d'amplifier des dérégulations dans les processus de défense immunitaire. Certains des motifs des domaines *gag*, *env* et de type LTR peuvent être associés à une fonction particulière ou peuvent contribuer à la fonction normale ou pathologique des domaines flanquants. Des recombinaisons avec un élément d'origine exogène, rétroviral ou non, peut donner lieu à la production de motifs nucléiques ou protéiques qui pourraient soit protéger, soit déclencher, ou favoriser ou aggraver une pathologie. De même, une structure rétrovirale contenant des éléments rétroviraux endogènes selon l'invention seraient à même de provoquer un processus pathologique après passage par un cycle transitoire exogène puis réintégration dans une région sensible ou critique du génome humain.

De même, la combinaison de motifs appartenant à la famille de HERV-7q, ou d'éléments induits par des motifs appartenant à la famille de HERV-7q, avec des motifs d'origine ou induits de manière exogène seraient à même de pouvoir déclencher, ou aggraver un processus pathologique ou au contraire de favoriser une protection ou une rémission partielle ou une guérison totale et définitive.

La détection rendue possible des domaines de type HERV-7q, suggère des applications possibles à la fois au niveau prophylactique, du pronostic et du diagnostic : par exemple des approches immunologiques ou d'amplification génique permettant de comparer des individus normaux servant de référence avec des patients, seraient à même de favoriser le dépistage, d'améliorer la détection précoce de la déclaration de la maladie et/ou de suivre l'évolution d'une pathologie chez des patients pouvant présenter une susceptibilité ou ayant déclaré la maladie ou encore chez des individus considérés comme normaux, selon les critères cliniques actuels.

Les sondes nucléiques et immunologiques spécifiques, telles que définies, dans la présente invention sont à même de favoriser l'identification et la détection de motifs anormalement exprimés dans le cadre de pathologies associées au cancer, ou de neuropathologies en particulier autoimmunes, au premier rang desquelles la sclérose en plaques.

Des manœuvres thérapeutiques peuvent être envisagées par usage de certaines des séquences nucléiques contenues dans HERV-7q et les séquences de la

même famille ou des structures polypeptidiques déduites ou par utilisation de peptides ou protéines, ou d'anticorps spécifiques.

La présente invention a également pour objet des séquences nucléiques hybrides, caractérisées en ce qu'elles comprennent des séquences ou motifs appartenant à la famille de HERV-7q, ou d'éléments induits par des motifs appartenant à la famille de HERV-7q, avec des motifs d'origine ou induits de manière exogène (séquences rétrovirales exogènes); de telles séquences hybrides sont vraisemblablement à même de pouvoir déclencher, ou aggraver un processus pathologique ou au contraire de favoriser une protection ou une rémission partielle ou une guérison totale et définitive.

La présente invention a également pour objet un réactif de diagnostic pour la détection différentielle de séquences nucléiques endogènes humaines complètes ou partielles, présentant des motifs rétroviraux, sélectionnés parmi les séquences SEQ ID NO :1 et/ou SEQ ID NO :2, caractérisé en ce qu'il est sélectionné dans le groupe constitué par les séquences SEQ ID NO :1-50, les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes, par les fragments nucléotidiques capables de définir ou d'identifier les séquences SEQ ID NO:1 et/ou SEQ ID NO:2 et toute séquence flanquante ou les chevauchants ainsi que par les fragments issus des régions codantes des séquences SEQ ID NO:1-24, correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires, éventuellement marquées avec un marqueur approprié.

Les séquences des sondes nucléiques, ribonucléiques et oligonucléotidiques utilisées seront choisies dans les régions *env* et *gag* ou leur régions flanquantes : par exemple les oligonucléotides amorces pour HERV-7q, seront choisis dans les régions situées entre les nucléotides 3065 et 4390 et les nucléotides 6965 et 9550, ainsi que dans toute séquence adjacente (amont ou aval) capable de permettre une amplification spécifique (figure 1).

Parmi les marqueurs appropriés, on peut citer, les isotopes radioactifs, les enzymes, les fluorochromes, des marqueurs chimiques (biotine), les haptènes (digoxigénine) et les anticorps ou analogues de bases appropriées.

De manière préférée :

- ledit réactif est sélectionné parmi les séquences SEQ ID NO:30-50 et est apte à être utilisé comme amorce.

- ledit réactif est sélectionné parmi les séquences suivantes :

5 un fragment de 1505 nt amplifié par le couple d'amorces SEQ ID NO:30 et SEQ ID NO:31 (amorces G1F et G1R),

un fragment de 2529 nt amplifié par le couple d'amorces SEQ ID NO:38 et SEQ ID NO:39 (amorces E1F et E1R) et est apte à être utilisé comme sonde.

10 La présente invention a également pour objet un procédé de détection rapide et différentiel des séquences nucléiques rétrovirales endogènes de type *env* ou *env* et *gag*, de leurs variants normaux ou pathologiques, par hybridation et/ou amplification génique, réalisé à partir d'un échantillon biologique, lequel procédé est caractérisé en ce qu'il comprend :

15 (a) une étape dans laquelle l'on met en contact un échantillon biologique à analyser avec au moins une sonde telle que définie ci-dessus et

(b) une étape dans laquelle on détecte par tout moyen approprié, le ou les produits résultant de l'interaction séquence nucléotidique-sonde.

Conformément audit procédé, il peut comprendre :

20 * préalablement à l'étape (a) :

. une étape de préparation du tissu ou du liquide biologique concerné,

. une étape d'extraction de l'acide nucléique à détecter, et

. au moins un cycle d'amplification génique et

25 * postérieurement à l'étape (b) :

. une étape de comparaison des séquences nucléiques obtenues dans ledit échantillon biologique avec les séquences rétrovirales endogènes humaines selon l'invention par tout moyen approprié et notamment par séquençage, Southern-blot, coupure de restriction, SSCP ou toute autre méthode permettant d'identifier une insertion ou une délétion ou encore une simple mutation entre les différentes séquences comparées.

30

Conformément à l'invention, les séquences rétrovirales endogènes humaines selon l'invention sont ainsi comparées aux séquences nucléiques présentes dans l'échantillon biologique à analyser et permettent la détection de séquences homologues de patients atteints de pathologies, susceptibles de mettre en jeu une
 5 modification de leur génome.

De manière avantageuse, lesdites comparaisons géniques sont menées à partir d'ADN génomique provenant d'individus témoins et de patients.

Une amplification génique classique par PCR sera menée à l'aide d'amorces 5' -sens et 3' -antisens encadrant ou comprenant la zone à étudier (zone *env*
 10 ou zone *gag*).

Également de manière avantageuse, les séquences des sondes nucléiques, ribonucléiques et oligonucléotidiques utilisées sont choisies dans les régions *env* et *gag* ou leurs régions flanquantes : par exemple les oligonucléotides amorces pour HERV-7q, seront choisis dans les régions situées entre les nucléotides
 15 3065 et 4390 et les nucléotides 6965 et 9550, ainsi que dans toute séquence adjacente (amont ou aval) capable de permettre une amplification spécifique (figure 1), comme précisé ci-dessus. Elles sont de préférence sélectionnées dans le groupe constitué par

un fragment de 1505 nt amplifié par le couple d'amorces SEQ ID NO:30 et SEQ ID NO:31 (amorces G1F et G1R),

20 un fragment de 2529 nt amplifié par le couple d'amorces SEQ ID NO:38 et SEQ ID NO:39 (amorces E1F et E1R).

L'étape d'amplification génique est notamment réalisée à l'aide d'une des techniques d'amplification génique suivante : amplification par la Q β -réplicase, PCR, LCR, ERA, CPR ou SDA.

25 La présente invention a également pour objet un procédé de détection des transcrits, tels que définis ci-dessus, caractérisé en ce qu'il comprend :

- le prélèvement des ARN messagers provenant de tissus témoins et de tissus prélevé chez des patients et

- l'analyse qualitative et/ou quantitative desdits ARNm, par hybridation *in situ*, par dot-blot, Northern-blot, RNase mapping ou RT-PCR, à l'aide d'un
 30 réactif de diagnostic tel que défini ci-dessus.

La présente invention a également pour objet des produits de traduction, caractérisés en ce qu'ils sont codés par une séquence nucléotidique telle que définie ci-dessus.

5 La présente invention a également pour objet un peptide, caractérisé en ce qu'il est susceptible d'être exprimé à l'aide d'une séquence nucléotidique sélectionnée dans le groupe constitué par les séquences SEQ ID NO:1-24, telles que définies ci-dessus.

Ledit peptide englobe également les peptides ou polypeptides dérivés comprenant entre 5 et 540 aminoacides (SEQ ID NO:25-29 et SEQ ID NO:51 et
10 leurs fragments d'au moins 5 aminoacides).

Lesdits peptides sont traduits à partir des séquences nucléiques telles que définies ci-dessus, selon les combinaisons offertes par l'usage des différents cadres de lecture possibles.

Selon un mode de réalisation avantageux desdits peptides, ils sont
15 notamment sélectionnés parmi les séquences SEQ ID NO:25-29 et SEQ ID NO :51

Selon un autre mode de réalisation avantageux desdits peptides, ils sont obtenus à partir des séquences nucléiques telles que définies ci-dessus, dans lesquelles au moins un codon non-sens peut être remplacé par un codon codant pour l'un des aminoacides suivants : Phe (F), Leu (L), Ser (S), Tyr (Y), Cys (C), Trp (W),
20 Gln (Q), Arg (R), Lys (K), Glu (E) ou Gly (G).

L'invention englobe ainsi les peptides déduits ou les protéines déduites correspondant à tout ou partie des séquences nucléiques décrites dans l'invention, et présentant éventuellement des modifications avec les séquences de références décrites dans l'invention, lorsqu'ils sont exprimés chez certains patients. En
25 particulier, l'invention englobe les séquences complètes ou partielles obtenues selon les 3 cadres de lecture sens et les 3 cadres de lecture inverses et complémentaires. (voir SEQ ID NO:22-24)

De manière avantageuse, la protéine env de HERV-7q selon l'invention présente :

30 - des sites de N-glycosylation. La glycosylation des protéines d'enveloppe des rétrovirus semble être directement associée à leurs propriétés fonc-

tionnelles, par exemple en influençant le nombre des déterminants disponibles dans les cellules-T ou en favorisant la reconnaissance des antigènes par les cellules-T. La glycosylation pourrait jouer un rôle dans la déclaration ou l'extension d'une pathologie à incidence autoimmune. Les glycosylations sont nécessaires au maintien de la conformation de certains épitopes, en particulier lors de la réalisation d'une protéine d'enveloppe recombinante à fin de mise au point d'un réactif de diagnostic et pour favoriser l'efficacité d'un éventuel vaccin. Positions 171, 210, 216, 236, 244, 283 et 411. Nombre prévu au hasard : 3.2

- des sites de prénylation. La prénylation est un mécanisme essentiel de la fixation à la membrane cellulaire et pour le ciblage de certaines protéines. Ce processus de ciblage pourrait être essentiel pour l'élaboration d'agents thérapeutiques spécifiques aptes à interférer dans la réalisation et la régulation du trafic de complexes cellulaires mettant en jeu des protéines impliquées dans les interactions, la croissance et les mouvements cellulaires. Positions 188 et 290. Nombre prévu au hasard : 1.8

- des sites de ciblage dans le réticulum endoplasmique. Ces sites permettraient d'assurer le ciblage vers le réticulum endoplasmique afin d'effectuer les modifications nécessaires pour favoriser le franchissement membranaire. Positions 353 et 431. Nombre prévu au hasard : 0.2

Lesdits peptides ou protéines peuvent présenter avantageusement des propriétés biologiques.

Les produits protéiques générés par les séquences rétrovirales endogènes ou produits parallèlement peuvent avantageusement être caractérisés par des micro-méthodes d'analyse et de quantification des peptides et des protéines: HPLC/FPLC ou équivalent, électrophorèse capillaire ou équivalent, techniques de microséquençages (méthode d'Edman ou équivalent, spectrométrie de masse...).

L'invention a également pour objet des anticorps dirigés contre l'un ou plusieurs des peptides décrits ci-dessus et leur utilisation pour la mise en œuvre d'une méthode de détection *in vitro*, notamment différentielle de la présence d'une telle séquence chez un individu.

Lesdits anticorps sont avantageusement des anticorps polyclonaux ou monoclonaux obtenus par une réaction immunologique d'un organisme humain,

mammifères, oiseaux ou autres espèces vis-à-vis des protéines, telles que définies ci-dessus.

La présente invention a pour objet un procédé de dépistage immunologique différentiel de séquences rétrovirales endogènes humaines de la famille HERV-7q normales ou pathologiques, caractérisé en ce qu'il comprend la mise en contact d'un échantillon biologique avec un anticorps selon l'invention, la lecture du résultat étant révélée par un moyen approprié, notamment EIA, ELISA, RIA, fluorescence.

A titre d'illustration, une telle méthode de diagnostic *in vitro* selon l'invention comprend la mise en contact d'un échantillon biologique prélevé chez un patient, avec des anticorps selon l'invention et la détection à l'aide de tout procédé approprié, notamment à l'aide d'anti-immunoglobulines marquée, des complexes immunologiques formés entre les protéines produites normalement ou pathologiquement et les anticorps.

Des anticorps monoclonaux ou polyclonaux, produits à partir d'antigènes correspondants à des peptides de synthèse, de polypeptide ou protéines recombinants, permettent de suivre l'expression des peptides ou protéines produits normalement ou pathologiquement. L'analyse est de préférence effectuée par ELISA, ou équivalent, Western-blot ou équivalent, ou par immunohistochimie.

Les peptides ou protéines, issus des séquences rétrovirales endogènes ou dont l'expression est associée à l'expression de ces séquences rétrovirales endogènes, sont recherchés et identifiés.

La présente invention a également pour objet un procédé d'identification et de détection de motifs rétroviraux endogènes, anormalement exprimés dans le cadre de pathologies associées au cancer, ou de neuropathologies en particulier autoimmunes, au premier rang desquelles la sclérose en plaques, caractérisé en ce qu'il comprend l'analyse comparée des séquences extraites d'un échantillon biologique avec les séquences selon l'invention.

La présente invention a également pour objet l'application des séquences nucléiques ou des séquences protéiques selon l'invention au diagnostic, au pronostic, à l'évaluation de la susceptibilité génétique, à toutes maladies humaines

induites, innées ou acquises en particulier celles à composantes cancéreuses, auto-immunes et/ou à incidence neurologique, comme la sclérose en plaques, les syndromes associés et les maladies neurodégénératives où intervient tout ou partie des séquences nucléiques selon l'invention et des formes endogènes ou exogènes apparentées.

La présente invention a également pour objet des séquences nucléiques hybrides, caractérisées en ce qu'elles comprennent des séquences ou motifs nucléiques selon l'invention, combinés avec des séquences ou motifs d'origine endogène ou d'origine ou induits de manière exogène.

La présente invention a, en outre, pour objet un vecteur recombinant de clonage ou d'expression, caractérisé en ce qu'il comprend une séquence nucléique conforme à l'invention.

Outre les dispositions qui précèdent, l'invention comprend encore d'autres dispositions, qui ressortiront de la description qui va suivre, qui se réfère à des exemples de mise en œuvre du procédé objet de la présente invention ainsi qu'aux dessins annexés, dans lesquels :

- Figure 1. Séquence nucléique humaine HERV-7q, dont l'analyse et le traitement permettent de caractériser une nouvelle structure rétrovirale endogène. Les régions nucléiques répétées de type R1 et R2 et les domaines *gag*, *pol* et *env* sont soulignés. Les domaines de type *gag* et *env* sont en italiques. La région homologue à une partie 3' non-codante de Rab7 est doublement soulignée.

- Figure 2. Cartographie de la région rétrovirale endogène humaine HERV-7q. La partie haute de la figure correspond à une région anonyme du génome humain située sur le bras long du chromosome 7. On peut identifier les domaines répétés (1), *gag* (2), *pol* (3) et *env* (4) de HERV-7q. La région *env* C-terminale (4.3) se prolonge en amont en un long cadre de lecture ouvert (4.2). Le domaine 4.1, correspond à la région N-terminale du domaine *env*.

- Figure 3. Comparaison des séquences nucléiques répétées situées aux bornes de HERV-7q. Les régions nucléiques répétées 5'(haut) et 3'(bas), sont comparées et les bases identiques sont indiquées par deux points.

- Figure 4. Séquence déduite présentant un cadre de lecture ouvert, dans le domaine de type-env de HERV-7q selon la règle du plus long cadre de lecture ouvert.

5 - Figure 5. Séquences autour du domaine CKS-17 identifiées dans différents domaines *env* déduits de la famille de HERV-7q et comparaison avec des motifs CKS-17 de référence.

1) HE2 - 2) HERV-7q - 3) N° d'accès à GenBank: M85205 - 4) HE7 - 5) HE9 - 6) CKS-17: le motif peptidique doué de propriétés immunomodulatrices est souligné - 7) gp20 de rétrovirus de type-D (SRV-Pc).

10 - Figure 6. Séquence déduite possible du domaine de type-gag identifié dans HERV-7q établie selon la règle du plus long cadre de lecture ouvert. X et / correspondent respectivement à un codon non-sens et à un décalage de cadre de lecture. La séquence soulignée correspond au début du domaine *pol*.

15 - Figure 7. Comparaison des régions nucléiques couvrant la région *gag* de HERV-7q (haut) et HERV-TcR (bas) et leurs régions flanquantes. Les bases identiques sont spécifiées par deux points.

20 - Figure 8. Exemple d'alignements nucléiques du domaine de type *env* de HERV-7q avec des domaines de type *env* similaires présents dans des séquences rétrovirales endogènes humaines de la même famille. Les codons non sens sont soulignés : 1) HERV-7q - 2) HE2 - 03) HE3 - 04) HE4.

25 - Figure 9. Alignements nucléiques entre le domaine *gag* de HERV-7q et les domaines correspondants appartenant à la même famille. Comparaison avec des fragments de domaines *gag* isolés d'agents rétroviraux infectieux. Séquences d'origine rétrovirale infectieuse: N° d'accession dans la banque de données EMBL : 1) A60168 - 2) A60201 - 3) A60200 - 4) A60171. Séquences rétrovirales endogènes humaines: 5) HERV-7q - 6) HG11 - 7) HG3. Les chiffres indiqués dans les séquences endogènes, correspondent au nombre de nucléotides insérés afin d'optimiser l'alignement avec les séquences de type *gag* identifiées dans des rétrovirus d'origine infectieuse.

30 - Figure 10. Alignement d'un motif *gag* protéique déduit (haut) appartenant à un rétrovirus infectieux (N° d'accession EMBL : A60200) avec le motif

gag protéique déduit (bas) identifié dans HERV-7q. Les codons non-sens sont en gras et soulignés. Les acides aminés identiques sont spécifiés par 2 tirets. Un tiret indique une délétion ou un acide aminé homologue.

5 - Figure 11. Alignement d'un motif *env* (haut) appartenant à un rétrovirus infectieux (N° d'accèsion EMBL : A60170) avec le motif *env* (bas) identifié dans HERV-7q. Les nucléotides homologues sont spécifiés par deux points et les délétions par un tiret.

10 - Figure 12. Comparaison entre le domaine *env* de HERV-7q (haut) et le domaine *env* de HERV-9 (bas). L'homologie de 66 % se limite à la région 3' du domaine *env* de HERV-7q et HERV-9, respectivement entre les nucléotides 8976 nt et 9500 nt de HERV-7q et les nucléotides 2898 nt et 3465 nt de HERV-9 (N° d'accèsion à GenBank : X57147). De nombreuses insertions/délétions sont aussi observées.

15 - Figure 13. Comparaison entre les domaines de type *env*, de HERV-7q et d'une séquence rétrovirale infectieuse exogène (n° d'accèsion EMBL : A60170).

Il doit être bien entendu, toutefois, que ces exemples sont donnés uniquement à titre d'illustration de l'objet de l'invention, dont ils ne constituent en aucune manière une limitation.

20 **EXEMPLE 1** : Détection, par amplification génique, d'une séquence nucléique appartenant à un domaine de type *gag* ou *env* selon l'invention, dans un échantillon d'ADN génomique d'origine humaine ou de mammifères.

L'amplification génique s'effectue à partir d'ADN génomique isolé à partir du sang. Un traitement anticoagulant est effectué avec 1 ml d'une solution de
25 citrate (pour un litre : 4,8 g de d'acide citrique, 13,2 g de citrate de sodium, 14,7 g de glucose) pour 6 ml de sang frais. Après centrifugation de 20 ml de sang pendant 15 mn à 13.0000 g, le surnageant est éliminé et la fraction enrichie en globules blancs est transférée dans un nouveau tube, puis recentrifugée dans les mêmes conditions que précédemment. La fraction enrichie en globules blancs est resuspendue dans un
30 tampon d'extraction (10 mM Tris-HCl, 0,1 M EDTA, 20 µg/ml de RNase pancréatique traitée afin d'éliminer les DNases, 0,5 % SDS, pH 8,0), puis incubée pendant 1 heure

à 37°C. La protéinase K est ajoutée à une concentration finale de 100 µg/ml. La suspension des cellules lysées est incubée à 50°C durant 3 heures sous agitation périodique, puis traitée par un volume égal de phénol équilibré par du Tris-HCl 0,5 M, pH 8,0. L'émulsion formée est placée sur une roue pendant une heure, puis centrifugée à 5000 g pendant 15 mn à température ambiante. La solution aqueuse est traitée déprotéinisée par une triple extraction phénolique afin d'obtenir un niveau de purification correspondant à un rapport final d'absorbance A260/A280 supérieur à 1,75. La fraction aqueuse est précipitée par 0,2 vol. d'acétate de sodium 10 M et 2 vol. d'éthanol. L'ADN est alors soit prélevé avec l'extrémité d'une pipette pasteur recourbée, soit centrifugé à 5000 g pendant 5 mn à température ambiante. L'ADN ou le culot d'ADN est lavé deux fois par de l'éthanol à 70 %, puis repris dans 1 ml de TE pH 8,0 afin d'être élué sous agitation douce pendant 12 à 24 heures.

Des oligonucléotides spécifiques des séquences endogènes décrites selon l'invention sont choisis pour amplifier la région *gag* ou *env* des régions rétrovirales endogènes décrites selon l'invention. L'ADN génomique étudié provient de patients présentant des pathologies comme la sclérose en plaques et d'individus réputés sains.

Les ADN polymérases thermostables utilisées ont été choisies pour leur grande fidélité lors du processus d'amplification, comme la Vent, ADN polymérase (Biolabs) ou équivalent, et sont utilisées selon les conditions préconisées par le fournisseur.

La stratégie d'amplification utilise selon les cas une simple PCR, ou une PCR nichée ou semi-nichée.

Oligonucléotides utilisés pour amplifier la région *gag* :

- amorce G1F, sens, localisée dans la région amont du domaine *gag* de *HERV-7q* (SEQ ID NO:30),
- amorce G1R, anti-sens, localisée dans la région 3' terminale du domaine *gag* (SEQ ID NO:31),

Le fragment de 1505 nt amplifié par le couple G1F-G1R : 1505 nt est utilisé afin de générer les sondes aptes à hybrider les différents produits d'amplification des PCR.

- amorce G2F, sens nichée (SEQ ID NO:32),
- amorce G2R, anti-sens nichée (SEQ ID NO:33),
- amorce G4F, sens nichée (SEQ ID NO:34),
- amorce G3F, sens nichée (SEQ ID NO:35),
- amorce G4R, anti-sens nichée (SEQ ID NO:36),
- amorce G5R, anti-sens nichée (SEQ ID NO:37),

Oligonucléotides utilisés pour amplifier la région *env* de HERV-7q :

- amorce E1F, sens (SEQ ID NO:38),
- amorce E1R, anti-sens (SEQ ID NO:39),

Le fragment de 2529 nt amplifié par le couple d'amorces E1F-E1R, est utilisé afin de générer les sondes aptes à hybrider les différents produits d'amplification des PCR.

- amorce E2F, sens (SEQ ID NO:40),
- amorce E2R, antisens (SEQ ID NO:41),
- amorce E3F, sens (SEQ ID NO:42),
- amorce E3R, anti-sens (SEQ ID NO:43),
- amorce E4F, sens (SEQ ID NO:44),
- amorce E4R, anti-sens (SEQ ID NO:45),
- amorce E5F, sens (SEQ ID NO:46),
- amorce E6F, sens (SEQ ID NO:47)
- amorce E5R (SEQ ID NO:48).
- amorce ExF (SEQ ID NO:49)
- amorce ExR (SEQ ID NO:50)

La PCR est réalisée à partir de 50 à 200 ng d'ADN génomique. Les conditions de PCR sont celles préconisées par le fournisseur. Les conditions cycliques d'amplification sont réalisées dans 50 µl : une dénaturation de 94°C pendant 1 min., une hybridation de 70°C pendant 1 min., et une élongation à 72 °C pendant 1 à 2 min., selon les fragments amplifiés. Après 35 cycles, une réaction terminale est menée à 72°C pendant 10 min. Le séquençage automatique des échantillons amplifiés est réalisé à l'aide d'un séquenceur Applied Biosystems de type ABI 377 ou autre modèle comparable, selon les protocoles fournis par le constructeur.

Dans le cas d'une PCR nichée ou semi-nichée, les mêmes conditions expérimentales sont utilisées, à la seule différence que l'échantillon d'ADN génomique est remplacé par 5 à 10 µl du produit d'amplification issu de la première PCR.

Deux amplifications indépendantes sont réalisées à partir du même échantillon. Une réaction de contrôle est réalisée en remplaçant l'échantillon d'ADN par de l'eau afin de détecter d'éventuels contaminants.

EXEMPLE 2 : Détection par amplification génique d'une séquence nucléaire selon l'invention dans un échantillon biologique d'ADN génomique prélevés chez des patients présentant une pathologie candidate déclarée ou la suspicion de cette pathologie.

Le protocole d'amplification est le même que dans l'exemple 2, mis à part l'origine de l'échantillon qui provient de patients présentant une pathologie candidate. Un échantillon d'ADN génomique réputé normal est systématiquement intégré dans l'ensemble des échantillons pathologiques amplifiés puis analysés.

Les produits de PCR sont séparés sur un gel d'agarose à 1,5 %, puis transférés en présence de soude 0,4 N sur une membrane de nylon chargé. Une hybridation est réalisée avec une sonde spécifique correspondant aux fragments de PCR amplifiés soit par les couples G1F-G1R soit par le couple E1F-E1R. La sonde est marquée par incorporation de dUTP-digoxygénine selon le protocole du fournisseur (Boehringer Mannheim). L'hybridation est effectuée dans un tampon d'hybridation (5XSSC, 50 % formamide, 0,1 % lauroyl-sarcosine, 0,02 % SDS, 2 % de réactif de blocage Boehringer) pendant une nuit à 42°C. Le Southern est lavé 2 fois 5 min. à température ambiante dans une solution de 2XSSC, 0,1% SDS. Puis un lavage à haute stringence est effectué à deux reprises pendant 15 min. à 55°C dans une solution 0,1XSSC, 0,1 % SDS. L'hybridation est révélée selon le protocole du fournisseur (Boehringer Mannheim), en présence d'un substrat chimioluminescent de la phosphatase alcaline, de type CSPD ou CDP-STAR. Le filtre est révélé après une exposition de 15min. à 60 min.

Une analyse par SSCP (« *single strand conformation polymorphism* ») permet de détecter des modifications discrètes de la séquence des fragments amplifiés par PCR. La PCR est menée en présence de dCTP marqués au P³². L'échan-

tillon a analyser est dénaturé à 95°C pendant 10 min., en présence de tampon de charge, puis immédiatement chargé sur un gel de polyacrylamide à 10%, contenant 7.5% de glycérol. La migration s'effectue à 4°C à 8-10 W. Le gel est séché puis autoradiographié.

- 5 Les fragments de PCR susceptibles de présenter une altération de leur séquence nucléotidique sont séquencés selon l'exemple 2.

Une hybridation à l'aide d'un oligonucléotide spécifique (17 mers à 20 mers) correspondant à la région nucléotidique modifiée permet d'identifier les échantillons présentant une modification identique (méthode ASO). Brièvement le
10 southern est hybridé avec un oligonucléotide marqué distalement soit au P³², soit en présence de digoxygénine (selon le protocole de Boehringer Mannheim) puis lavé dans des conditions stringentes à 65°C dans une solution 6XSSC, 0.05% pyrophosphate de sodium.

EXEMPLE 3 : Détection d'une protéine selon l'invention dans un échantillon
15 **biologique.**

- Préparation d'une fraction protéique purifiée de liquide céphalo-rachidien de patients atteints de SEP

Après un traitement à 56°C pendant 30 min, et élimination des immunoglobulines sur une colonne de protéine G HiTrap (Pharmacia), l'équivalent de
20 10 ml de LCR est déposé sur une colonne de DEAE Sepharose CL-6B (Pharmacia). L'élution est réalisée en Tris-HCl 20 mM pH 8,8, et un gradient de 0 à 0,4 M de NaCl, puis la fraction est dialysée 2 fois contre du tampon phosphate-NaCl (PBS). Après concentration sur Ultrafree-MC (Millipore), la fraction est déposée sur une colonne de Superose 12 (FPLC Pharmacia) et éluée en présence de PBS. Après séparation par
25 électrophorèse en gel de polyacrylamide-SDS, et électro-transfert sur une membrane d'Immobilon-P (Millipore), les bandes protéiques sont soumises à une hydrolyse trypsique ménagée.

- Analyse de la fraction protéique par spectrométrie de masse

Les peptides digérés en présence de trypsine, sont analysés par la
30 méthode de MALDI-TOF, qui permet l'analyse de peptides présents en mélange. (COTTRELL J.S., Pept. Res., 1997, 7, 115-124). Les peptides caractérisés en fonction

de leur masse sont comparés aux protéines et aux protéines associées selon l'invention.

EXEMPLE 4 : Détection d'anticorps spécifiques anti-domaine *env* de HERV-7q.

L'identification d'un long cadre de lecture ouvert au sein de la
 5 séquence *env* de HERV-7q, a permis de déterminer une séquence protéique déduite
 SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29
 d'une région dudit gène, référencée par SEQ ID NO:22.

Les séquences de protéines déduites des séquences ID NO:23, 25,
 27, 28 et 29 sont positionnées comme suit par rapport à la figure 1 ou à la séquence ID
 10 NO:3 :

SEQ ID NO:23 : début de la séquence codante : position 7874, fin de
 la séquence codante 1^{er} codon non-sens (position 9493)

SEQ ID NO:25 : début de la séquence codante : position 7874, fin de
 la séquence codante 1^{er} codon non-sens (position 9493) (cadre de lecture 1)

15 SEQ ID NO:27 : début de la séquence codante : position 6970, fin de
 la séquence codante 1^{er} codon non-sens (position 9493) (cadre de lecture 1)

SEQ ID NO:28 : début de la séquence codante : position 6971, la fin
 du cadre de lecture est décalée selon le cas de 1, 2 ou 3 codons

SEQ ID NO:29 : début de la séquence codante : position 6972, la fin
 20 du cadre de lecture est décalée selon le cas de 1, 2 ou 3 codons

Différents peptides correspondant à tout ou partie des SEQ ID
 NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29 ont été
 synthétisés par génie génétique afin de tester leur spécificité antigénique vis à vis de
 séra ou de tissus de patients atteints de SEP, par exemple. Brièvement, tout ou partie
 25 de la région *env* de HERV-7q est sous clonée dans les vecteurs pQE30, 31 et 32. Les
 vecteurs pQE30, 31 et 32 contiennent en 5' du multi-site de clonage les séquences
 consensuelles pour la transcription (le promoteur fort du bactériophage T5, 2 opéra-
 teurs de l'opéron lactose), la traduction (un site d'accrochage ribosomal synthétique).
 De même, pQE30, 31 et 32 possèdent en 3', le terminateur de transcription du phage 1
 30 ainsi qu'un codon "Stop" pour la traduction. L'expression de la protéine s'effectue
 après transformation dans *E. coli* M15. Le plasmide pQE30, 31 et 32 possèdent en

amont du site de polyclonage la séquence codante pour une suite de 6 histidines présentant une affinité pour les ions nickel. Cet enchaînement permet la purification de la protéine chimérique exprimée, par adsorption sur une résine constituée d'un ligand chélatant, l'acide nitrilotriacétique (NTA), chargé de 4 ions nickel (résine NI-
 5 NTA, Qiagen).

La transformation s'effectue par électroporation ou traitement au chlorure de calcium. Par exemple, une colonie d'*E. coli* M15 est incubée dans 100 ml de milieu LB contenant 250 µg de kanamycine, sous agitation à 37°C jusqu'à l'obtention d'une DO⁶⁰⁰ de 0,5. Après une centrifugation de 5 minutes à 2000g à 4°C,
 10 le culot bactérien est repris dans 30 ml de solution TFB1 (100 mM de chlorure de rubidium, 50 mM de chlorure de manganèse, 30 mM d'acétate de potassium, 10 mM CaCl₂, 15% glycérol, pH 5.8), à 4°C pendant 90 minutes. Après une centrifugation de 5 minutes à 2000g à 4°C, le culot bactérien est repris dans 4 ml de solution TFB2 (10
 mM de chlorure de rubidium, 10 mM de MOPS, 75 mM CaCl₂, 15% de glycérol, pH
 15 8). Les cellules peuvent être gardées à -70°C par aliquot de 500 µl. 20 µl de la ligation et 125 µl de cellules compétentes sont mélangés et placés dans la glace 20 minutes. Après un choc thermique de 42°C pendant 90 secondes, les cellules sont agitées 90 minutes à 37°C dans 500 ml de milieu Psi-broth (milieu LB complémenté par 4 mM de MgSO₄, 10mM de chlorure de potassium). Les cellules transformées sont étalées
 20 sur des boîtes LB-agar complémentées par 25 µg/ml de kanamycine, et 100µg/ml d'ampicilline, et les boîtes sont incubées une nuit à 37°C.

Les clones potentiellement recombinants sont repiqués de manière ordonnée sur un filtre de nylon déposé sur une boîte LB-agar complémentée par 25 µg/ml de kanamycine et 100 µg/ml d'ampicilline. Après une nuit à 37°C, les clones
 25 recombinants sont repérés par hybridation de l'ADN plasmidique avec la sonde nucléotidique amplifiée par PCR avec le couple d'amorces selon SEQ ID NO:38 et SEQ ID NO:39.

Une colonie indépendante, contenant l'insert, est inoculée à 20 ml de milieu LB complémentée par 25 µg/ml de kanamycine et 100 µg/ml d'ampicilline.
 30 Après une nuit à 37°C sous agitation, 500 ml de même milieu sont incubés au 1/50° par cette préculture jusqu'à l'obtention d'une DO⁶⁰⁰ de 0,8, puis 1 à 2 mM final d'IPTG

est ajouté. Après 5 heures, les cellules sont centrifugées 20 minutes à 4000 g.

Une partie du culot cellulaire est repris dans 5 ml de tampon de sonication (50 mM de phosphate de sodium pH 7,8, 300 mM NaCl) puis placé dans la glace. Après une rapide sonication, les cellules sont centrifugées 20 minutes à 10000 g. Une partie du culot cellulaire est repris dans 10 ml d'une solution 30 mM Tris/HCl-20% sucrose pH8. Les cellules sont incubées 5 à 10 minutes sous agitation, après adjonction de 1 mM EDTA. Après une centrifugation de 10 minutes à 8000 g à 4°C, le culot est repris dans 10 ml de 5 mM de MgSO₄ glacé. Après 10 minutes dans la glace sous agitation, les cellules sont centrifugées 10 minutes à 8000 g à 4°C.

Le culot est repris par 5 ml/g dans du tampon A (6 M GuHCl (chlorhydrate de guanidine), 0,1M phosphate de sodium, 0,01M Tris/HCl, pH 8), 1 heure à température ambiante. Le lysat est centrifugé 15 minutes à 10000 g à 4°C, et le surnageant est complété par 8 ml de résine Ni-NTA, prééquilibrée dans du tampon A. Après 45 minutes à température ambiante, la résine est coulée dans une colonne, lavée par 10 fois le volume de la colonne par du tampon A puis par 5 fois le volume de la colonne par du tampon B (8 M urée, 0,1 M phosphate de sodium, 0,01 M Tris/HCl, pH 8). La colonne est lavée par du tampon C (8 M urée, 0,1M phosphate de sodium, 0,01 M Tris/HCl, pH 6,3) jusqu'à ce que l'A280 soit inférieur à 0,01. La protéine recombinante est éluée par 10 à 20 ml de tampon D (8 M urée, 0,1 M phosphate de sodium, 0,01 M Tris/HCl, pH 5,9) puis par 10 à 20 ml de tampon E (8 M urée, 0,1 M phosphate de sodium, 0,01 M Tris/HCl, pH 4,5), puis par 20 ml de tampon F (6 M HCl, 0,2 M acide acétique). Après une analyse en SDS-PAGE, la ou les fractions purifiées contenant la protéine chimérique ont permis l'obtention d'anticorps chez le lapin. Les anticorps obtenus sont testés par Western-blot après révélation par un anticorps secondaire couplé à la phosphatase alcaline.

Des anticorps sont obtenus de la même manière, à partir de peptides synthétisés chimiquement selon la technique de Merrifield (G. Barany and B. Merrifield, 1980, dans *The peptides*, 2, 1-284, E. Gross et J. Meienhofer, Academic Press, New York).

Les anticorps spécifiques obtenus sont utilisés à fin de détection de l'expression sérique ou tissulaire de tout ou partie des séquences rétrovirales endo-

gènes selon l'invention, dans les cas normaux et pathologiques.

Les protéines d'origine sérique ou tissulaire, sont séparées sur gel d'acrylamide-SDS puis transférées sur un filtre de nitrocellulose à l'aide d'un appareil Novablot 2117-2250 (LKB). Le transfert est effectué sur une feuille de Hybond C-extra (Amersham) en utilisant un tampon CAPS 100 mM pH 11, méthanol, eau (V/V/V: 1/1/8) contenant 1 mM de CaCl_2 . Après un transfert de 1 heure à 0,8 mA/cm², la feuille est saturée une heure à température ambiante dans du PBS-0,5 % gélatine. La feuille est mise en présence de l'anticorps spécifique à la concentration de 1/1000 dans du PBS-0,25 % gélatine. Au bout de 2 heures, le filtre est lavé 3 fois 15 minutes dans du PBS-0,1 % de Tween-20, puis le filtre est incubé 30 minutes en présence d'un anticorps secondaire couplé à la phosphatase alcaline (Promega), dilué au 1/7500 dans du PBS-0,25% gélatine. Après trois lavages dans du PBS-0,1 % de Tween-20, le filtre est équilibré dans un tampon (100 mM de Tris-HCl pH 9,5, 100 mM de NaCl, 5 mM de MgCl_2). La révélation est effectuée en présence de 45 µl de NBT à 75 mg/ml et 35 µl de BCIP à 50 mg/ml, pour 10 ml de tampon de phosphatase alcaline.

Les protéines chimériques obtenues par génie génétique, sont utilisées aussi à fin de tests d'activité biologique, comme par exemple pour le test d'activité biologique du peptide de type CKS-17 identifié dans le domaine *env* de HERV-7q (figure 5).

EXEMPLE 5 : Obtention de sondes ribonucléiques codant pour les séquences *env* de HERV-7q.

Les fragments de PCR obtenus sont sous clonés dans le plasmide PGEM 4Z (Promega) qui possède de par et d'autre de son site de polyclonage, les séquences promotrices pour les ARN polymérase SP6 et T7.

La méthode de compétence utilisée est l'électroporation. Le plasmide et le fragment de PCR sont hybridés dans un rapport de 50 ng de vecteur (coupé à Sma I) pour 100 ng de fragment de PCR (rendu à bout franc par traitement par le fragment de Klenow de l'ADN polymérase). L'incubation a lieu une nuit à 22°C, dans le tampon de ligation (66 mM Tris-HCl pH 7,5, 5 mM MgCl_2 , 1 mM dithioerythritol,

1 mM ATP) en présence de 1u. de T4 ADN ligase puis est arrêtée par dénaturation 10 minutes à 65°C. Parallèlement, la souche d'*E. Coli* JM 105 est ensemencée une nuit à 37°C dans du milieu LB. Cette préculture est diluée au 1/500 et placée à 37°C jusqu'à une DO⁶⁰⁰ égale à 1. Pour la suite du mode opératoire les cellules seront toujours

5 conservées au froid. Après une centrifugation de 5 minutes à 3500 g à 4°C, le culot cellulaire est resuspendu dans 1/4 vol. d'eau glacée ultra-pure. Cette étape est répétée 5 à 6 fois. Puis le culot est resuspendu dans 1/4000 vol. d'eau; 10 % de glycérol stérile sont ajoutés permettant la conservation des cellules électrocompétentes, par aliquots de 10 µl à 20°C. A 50 µl de cellules électrocompétentes est ajouté 1 µl de la ligation ;

10 le tout est soumis à une décharge électrique de 12,5 kV/cm, appliquée pendant 5,8 ms. Les cellules sont rapidement remises en suspension dans le milieu SOC, incubées 1 heure à 37°C, puis étalées, en présence de 2% X-Gal dans du diméthylformamide, et 10 mM d'IPTG, sur une boîte de gélose LB-agar supplémentée en ampicilline (100 µg/ml). Après une nuit à 37°C, les clones blancs potentiellement recombinants, sont

15 repiqués de manière ordonnée sur une boîte LB/ampicilline et parallèlement sur un filtre de nylon déposé sur une boîte LB/ampicilline. Ces deux boîtes sont incubées une nuit à 37°C. Les clones recombinants sont alors repérés par hybridation avec une sonde nucléique amplifiée par PCR avec le couple d'amorces selon SEQ ID NO:38 et SEQ ID NO:39 et marquée à la digoxygénine.

20 Les clones recombinants sont cultivés dans 50 ml de milieu LB/ampicilline (100 µg/ml) en agitation pendant une nuit à 37°C. Après une centrifugation à 3500 g pendant 15 minutes à 4°C, le culot bactérien est repris dans 4ml de tampon P1 (50 mM Tris-HCl, 10mM EDTA, 400 µg/ml RNase A, pH 8) et 4ml de tampon P2 (200 mM NaOH, 1% SDS). Le mélange est incubé à température ambiante

25 pendant 5 minutes. Après adjonction de 4ml de tampon P3 (2,55 M d'acétate de potassium, pH 4,8) le mélange est centrifugé à 12000 g pendant 30 minutes à 4°C. Le surnageant est appliqué sur une colonne Qiagen-type 100, prééquilibrée avec 2 ml de tampon QBT (750 mM NaCl, 50 mM MOPS, 15% éthanol, pH 7), la colonne est lavée avec 2 fois 4ml de tampon QC (1M NaCl, 50 mM MOPS, 15 % éthanol, pH 7) et

30 l'ADN est élué avec 2ml de tampon QF (1,2 M NaCl, 50mM MOPS, 15 % éthanol, pH 8). L'ADN est précipité avec 0,8 vol. d'isopropanol, et centrifugé à 12000 g à 4°C

pendant 30 minutes. Le culot est lavé avec de l'éthanol à 70 % glacé, puis l'ADN plasmidique est repris par 2 fois 150 µl de tampon TE.

Les sondes ribonucléiques sont utilisées comme sondes spécifiques, en particulier pour la détection des transcrits exprimés par les séquences rétrovirales endogènes selon l'invention.

Bibliographie :

- Benit L. et al., 1997. Cloning of a new murine endogenous retrovirus MuERV-L, with strong similarity of the human HERV-L element and with a *gag* coding sequence closely related to the Fv1 restriction gene. *J. Virol.* 71, 5652-5657.
- 10 - Coffin J.M. 1985. Endogenous retrovirus. In: "RNA tumor viruses" (Weiss R.A., Varmus H.E., Teich N.M., and Coffin J.M. eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Conrad B., Weissmahr R.N., Boni J., Arcari R., Schupbach J., and Mach B. 1997. A human endogenous retroviral superantigen as candidate autoimmunogene in type 1
- 15 diabetes. *Cell* 90, 303-313.
- Covey S.N. 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.* 14, 623-633.
- Hertig C., Coupar B.E., Gould A.R., and Boyle D.B. 1997. Field and vaccine strains of
- 20 fowlpox virus carry integrated sequences from the avian retrovirus, reticuloendotheliosis virus. *Virology* 235, 367-376.
- Hohenadl C., Leib-Mösch C., Hehleemann R., and Erfle Y. 1996. Biological significance of human endogenous retroviral sequences. *J. Acqui. Imm. Def. Synd. Hum. Retrovir.* 13, S268-S273.
- 25 - Kulkoski J.K., Jones S., Katz R.A., Mack J.P.G., and Skalka A.M. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* 12, 2331-2338.
- La Mantia G. et al, N.A.R., 1991, 19, 7, 1513-1520
- 30 - Patience C., Wilkinson D.A., and Weiss R.A. 1997. Our retroviral heritage. *Trends Genet.* 13, 116-120.

- Pearson W.R. 1994. Using the FASTA program to search protein and DNA sequence databases. *Methods Mol. Biol.* 24, 307-331.
- Perron H., Garson J.A., Bedin F., Beseme F., Paranhos-Baccala G., Komurian-Pradel F., Mallet F., Tuke P.W., Voisset C., Blond J.L., Lalande B., Seigneurin J.M., Mandrand B.
- 5 and the Collaborative Research Group on Multiple Sclerosis. 1997. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci. USA* 94, 7583-7588.
- Tönjes R.R. et al., *J. AIDS and Hum. Retrovirol*, 1996, 13, S261-S267
- Vitelli R., Chiarillo M., Lattero D., Bruni C.B., and Bucci C. 1996. Molecular cloning
- 10 and expression analysis of the human Rab7 GTP-ase complementary deoxyribonucleic acid. *Biochem. Biophys. Res. Commun.* 229, 887-890.
- Weber L.T., Miller M., Jaskolski M., Leis J., Skalka M., and Wlodawer A. 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* 243, 928-931.
- 15 - Wilkinson D., Mager D.L., and Leong J.A.C. 1994. Endogenous human retroviruses. In: "The Retroviridae" (Levy, J.A. ed), Plenum Press New York, , Vol. 3, 465-535.
- Xiong Y., and Eickbush, T. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353-3362.

20 Ainsi que cela ressort de ce qui précède, l'invention ne se limite nullement à ceux de ses modes de mise en œuvre, de réalisation et d'application qui viennent d'être décrits de façon plus explicite ; elle en embrasse au contraire toutes les variantes qui peuvent venir à l'esprit du technicien en la matière, sans s'écarter du cadre, ni de la portée, de la présente invention.

LISTE DE SEQUENCES

(1) INFORMATIONS GENERALES:

(i) DEPOSANT:

- (A) NOM: INSTITUT NATIONAL DE LA RECHERCHE MEDICALE -
INSERM
(B) RUE: 101 RUE DE TOLBIAC
(C) VILLE: PARIS
(E) PAYS: FRANCE
(F) CODE POSTAL: 75654 CEDEX

(ii) TITRE DE L'INVENTION: FAMILLE DE SEQUENCES NUCLEIQUES ET DE SEQUENCES PROTEIQUES DEDUITES PRESENTANT DES MOTIFS RETROVIRAUX ENDOGENES HUMAINS ET LEURS APPLICATIONS.

(iii) NOMBRE DE SEQUENCES: 51

(iv) FORME DECHIFFRABLE PAR ORDINATEUR:

- (A) TYPE DE SUPPORT: Floppy disk
(B) ORDINATEUR: IBM PC compatible
(C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS
(D) LOGICIEL: PatentIn Release #1.0, Version #1.30 (OEB)

(2) INFORMATIONS POUR LA SEQ ID NO: 1: 7env

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 2599 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

ATCCCCCTGCC TTAATCGCCA AGCTCCTTCA GGAGAACAAA GAACAGGCCA TTACCCTGGA	60
GAAGACTGGC AACTGATTTT ACCCACAAGC CCAAACCTCA GGGATTTCAG TATCTACTAG	120
TCTGGGTAGA TACTTTCACG GGTGGGCAG AGGCCTTCCC CTGTAGGACA GAAAAGGCCC	180
AAGAGGTAAT AAAGGCACTA GTTCATGAAA TAATTCCCAG ATTCGGACTT CCCCAGAGCT	240
TACAGAGTGA CAATAGCCCT GCTTCCAGG CCACAGTAAC CCAGGGAGTA TCCCAGGCGT	300
TAGGTATACG ATATCACTTA CACTGCGCCT GAAGGCCACA GTCCTCAGGG AAGGTCGAGA	360
AAATGAATGA AACACTCAA GGACATCTAA AAAAGCAAAC CCAGGAAACC CACCTCACAT	420
GGCCTGCTCT GTTGCCCTATA GCCTTAAAAA GAATCTGCAA CTTTCCCCAA AAAGCAGGAC	480
TTAGCCCATA CGAAATGCTG TATGGAAGGC CTTTCATAAC CAATGACCTT GTGCTTGACC	540
CAAGACAGCC AACTTAGTTG CAGACATCAC CTCCTTAGCC AAATATCAAC AAGTTCTTAA	600
AACATTACAA GGAACCTATC CCTGAGAAGA GGGAAAAGAA CTATTCCACC CTTGTGACAT	660
GGFATTAGTC AAGTCCCTTC CCTCTAATTC CCCATCCCTA GATACATCCT GGAAGGACC	720
CTACCCAGTC ATTTTATCTA CCCCAACTGC GGTAAAGTG GCTGGAGTGG AGTCTTGGAT	780
ACATCACACT TGAGTCAAAT CCTGGATACT GCCAAAGGAA CCTGAAAATC CAGGAGACAA	840

CGCTAGCTAT TCCTGTGAAC CTCTAGAGGA TTTGCGCCTG CTCTTCAAAC AACACCAGG	900
AGGAAAGTAA CTAAATCAT AAATCCCCAT GGCCCTCCCT TATCATATTT TTCTCTTTAC	960
TGTTCTTTTA CCCTCTTTCA CTCTCACTGC ACCCCCTCCA TGCCGCTGTA TGACCAGTAG	1020
CTCCCTTAC CAAGAGTTTC TATGGAGAAT GCAGCGTCCC GGAAATATTG ATGCCCCATC	1080
GTATAGGAGT CTTTCTAAGG GAACCCCCAC CTTCACTGCC CACACCCATA TGCCCCGCAA	1140
CTGCTATCAC TCTGCCACTC TTTGCATGCA TGCAAATACT CATTATTGGA CAGGAAAAAT	1200
GATTAATCCT AGTTGTCTCG GAGGACTTGG AGTCACTGTC TGTTGGACTT ACTTCACCCA	1260
AACGGTATG TCTGATGGGG GTGGAGTTCA AGATCAGGCA AGAGAAAAAC ATGTAAAAGA	1320
AGTAATCTCC CAACTCACCC GGTACATGG CACCTCTAGC CCCTACAAAG GACTAGATCT	1380
CTCAAACTA CATGAAACCC TCCGTACCCA TACTCGCCTG GTAAGCCTAT TTAATACCAC	1440
CCTCACTGGG CTCCATGAGG TCTCGGCCCA AAACCCTACT AACTGTTGGA TATGCCTCCC	1500
CCTGAACTTC AGGCCATATG TTTCAATCCC TGTACCTGAA CAATGGAACA ACTTCAGCAC	1560
AGAAATAAAC ACCACTTCCG TTTTAGTAGG ACCTCTGTT TCCAATCTGG AAATAACCCA	1620
TACCTCAAAC CTCACCTGTG TAAATTTAG CAATACTACA TACACAACCA ACTCCCAATG	1680
CATCAGGTGG GTAACCTCTC CCACACAAAT AGTCTGCCTA CCCTCAGGAA TATTTTTTGT	1740
CTGTGGTACC TCAGCCTATC GTTGTTTGAA TGGCTCTCA GAATCTATGT GCTTCCTCTC	1800
ATTCTTAGTG CCCCCTATGA CCATCTACAC TGAACAAGAT TTATACAGTT ATGTCATATC	1860
TAAGCCCCGC AACAAAAGAG TACCCATTCT TCCTTTTGT ATAGGAGCAG GAGTGCTAGG	1920
TGCACTAGGT ACTGGCATTG GCGGTATCAC AACCTCTACT CAGTTCTACT ACAAACATC	1980
TCAAGAACTA AATGGGGACA TGGAACGGGT CGCCGACTCC CTGGTCACCT TGCAAGATCA	2040
ACTTAACCTC CTAGCAGCAG TAGTCCTTCA AAATCGAAGA GCTTTAGACT TGCTAACCGC	2100
TGAAAGAGGG GGAACCTGTT TATTTTTAGG GGAAGAATGC TGTTATTATG TTAATCAATC	2160
CGGAATCGTC ACTGAGAAAG TTAAAGAAAT TCGAGATCGA ATACAACGTA GAGCAGAGGA	2220
GCTTCGAAAC ACTGGACCCT GGGGCCTCCT CAGCCAATGG ATGCCCTGGA TTCTCCCCTT	2280
CTTAGGACCT CTAGCAGCTA TAATATTGCT ACTCCTCTTT GGACCCTGTA TCTTTAACCT	2340
CCTTGTTAAC TTTGTCTCTT CCAGAATCGA AGCTGTAAAA CTACAAATGG AGCCCAAGAT	2400
GCAGTCCAAG ACTAAGATCT ACCGCAGACC CCTGGACCGG CCTGCTAGCC CACGATCTGA	2460
TGTTAATGAC ATCAAAGGCA CCCCTCCTGA GGAAATCTCA GCTGCACAAC CTCTACTACG	2520
CCCCAATTCA GCAGGAAGCA GTTAGAGCGG TCTCGGCCAA CCTCCCCAAC AGCACTTAGG	2580
TTTCTCTGTT GAGATGGGG	2599

(2) INFORMATIONS POUR LA SEQ ID NO: 2: gag

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 1326 paires d bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

GCCGCCTGGC ACTCCTGAGG GAAGTATAAA TTATAACACC ATCTTACAGC TAGACCTCTT	60
TTGTAGAAAA GGCAAATGGA GTGAAGTGCC ATAAGTACAA ACTTTCTTTT CATTAAGAGA	120
CAACTCACAA TTATGTAAAA AGTGTGATTT ATGCCCTACA GGAAGCCTTC AGAGTCTACC	180
TCCCTATCCC AGCATCCCCG ACTCCTTCCC CAACTAATAA GGACCCCCCT TCAACCCAAA	240
TGGTCCAAAA GGAGATAGAC AAAAGGGTAA ACAGTGAACC AAAGAGTGCC AATATTCCCC	300
AATTATGACC CCTCCAAGCA GTGGGAGGAA GAGAATTCGG CCCAGCCAGA GTGCATGTGC	360
CTTTTTCTCT CCCAGACTTA AAGCAAATAA AACAGACTT AGGTAAATTC TCAGATAACC	420
CTGATGGCTA TATTGATGTT TTACAAGGGT TAGGACAATT CTTTGATCTG ACATGGAGAG	480
ATATAATGTC ACTGCTAAAT CAGACACTAA CCCCAAATGA GAGAAGTGCC ACCATAACTG	540
CAGCCTGAGA GTTTGGCGAT CTCTGGTATC TCAGTCAGGT CAATGATAGG ATGACAACAG	600
AGGAAAGAGA ATGATTCCCC ACAGGCCAGC AGGCAGTTCC CAGTCTAGAC CCTCATTGGG	660
ACACAGAATC AGAACATGGA GATTGGTGCT GCAGACATTT GCTAACTTGT GTGCTAGAAG	720
GACTAAGGAA AACTAGGAAG AAGTCTATGA ATTACTCAAT GATGTCCACC ATAACACAGG	780
GAAGGGAAGA AAATCCTACT GCCTTTCTGG AGAGACTAAG GGAGGCATTG AGGAAGCGTG	840
CCTCTCTGTC ACCTGACTCT TCTGAAGGCC AACTAATCTT AAAGCGTAAG TTTATCACTC	900
AGTCAGCTGC AGACATTAGA AAAAACTTC AAAAGTCTGC CGTAGGCCCG GAGCAAACT	960
TAGAAACCCT ATTGAACTTG GCAACCTCGG TTTTTTATAA TAGAGATCAG GAGGAGCAGG	1020
CGGAACAGGA CAAACGGGAT TAAAAAAG GCCACCGCTT TAGTCATGAC CCTCAGGCAA	1080
GTGGACTTTG GAGGCTCTGG AAAAGGGAAA AGCTGGGCAA ATTGAATGCC TAATAGGGCT	1140
TGCTTCCAGT GCGGTCTACA AGGACACTTT AAAAAAGATT GTCCAAGTAG AAGTAAGCCG	1200
CCCCCTCGTC CATGCCCTT ATTTCAAGGG AATCACTGGA AGGCCCACTG CCCAGGGGA	1260
CAAAGGTCCT CTGAGTCAGA AGCCACTAAC CAGATGATCC AGCAGCAGGA CTGAGGGTGC	1320
CTGGGG	1326

(2) INFORMATIONS POUR LA SEQ ID NO: 3: HERV-7q

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 10499 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

CCCTGGGGCG GGCTTCCTTT CTGGGATGAG GGCAAACGC CTGGAGATAC AGCAATTATC	60
TTGCAACTGA GAGACAGGAC TAGCTGGATT TCCTAGGCCG ACTAAGAATC CCTAAGCCTA	120

GCTGGGAAGG	TGACCACGTC	CACCTTTAAA	CACGGGGCTT	GCAACTTAGC	TCACACCTGA	180
CCAATCAGAG	AGCTCACTAA	AATGCTAATT	AGGCAAAGAC	AGGAGGTAAA	GAAATAGCCA	240
ATCATCTATT	GCCTGAGAGC	ACAGCAGGAG	GGACAACAAT	CGGGATATAA	ACCCAGGCAT	300
TCGAGCTGGC	AACAGCAGCC	CCCCTTTGGG	TCCCTTCCCT	TTGTATGGGA	GCTGTTTTCA	360
TGCTATTTCA	CTCTATTAAA	TCTTGCAACT	GCACTCTTCT	GGTCCATGTT	TCTTACGGCT	420
CGAGCTGAGC	TTTTGCTCAC	CGTCCACCAC	TGCTGTTTGC	CACCACCGCA	GACCTGCCGC	480
TGACTCCCAT	CCCTCTGGAT	CCTGCAGGGT	GTCCGCTGTG	CTCCTGATCC	AGCGAGGCGC	540
CCATTGCCGC	TCCCAATTGG	GCTAAAGGCT	TGCCATTGTT	CCTGCACGGC	TAAGTGCCTG	600
GGTTTGTCT	AATTGAGCTG	AACACTAGTC	ACTGGGTTC	ATGGTTCTCT	TCTGTGACCC	660
ACGGCTTCTA	ATAGAACTAT	AACACTTACC	ACATGGCCCA	AGATTCCATT	CCTTGGGAATC	720
CGTGAGGCCA	AGAACTCCAG	GTCAGAGAAT	ACGAGGCTTG	CCACCATCTT	GGAAGCGGCC	780
TGCTACCATC	TTGGAAGTGG	TTCACCACCA	TCTTGGGAGC	TCTGTGAGCA	AGGACCCCCC	840
GGTAACATTT	TGGCAACCAC	GAACGGACAT	CCAAAGTGGT	GAGTAATATT	GGACCACTTT	900
CACTTGCTAT	TCTGTCCTAT	CCTTCCTTAG	AATTGGAGGA	AAATACCGGG	CACTTGTCGG	960
CCAGTTAAAA	ACGATTAGTG	TGGCCACCGG	ACTTAAGACT	CAGGTGTGAG	GCTATCTGGG	1020
GAAGGGCTTT	CTAACAACCC	CCAACCCTTC	TGGGTTGGGG	ACTTGGTTTG	CCTCAAGCCA	1080
GCTTCCACTT	TCAGTTTTCT	TGGGGAAGCC	GAGGGCCGAC	TAGAGGCAGA	AAGCTGTCTG	1140
CCTGAACTCC	CGGCAGTAGC	CGTTGAGAT	CATGGTGTAG	CCAGAAGTCT	CAACAGTCGC	1200
CCATGCATGC	ACCCCTATCT	TTCCTTCTGA	CCCATACCTC	CTGGGTCCCA	ACCACAACCT	1260
TCTTCAAAGT	GTAGCCCCAA	AATTCTCCTT	ACCTCTGAAT	ATACTTCCTC	TGATCCCTGC	1320
CTCCTAGGTA	CTATTGGTTC	AGACTTCCAT	TTCCTCTAGC	AAGTTGTATC	TCCAAAGGGA	1380
TCTAAGGAAG	CTCTGCGCTG	CGTCCTTAGG	CACCTAGGCT	ATAACCCAGG	GAGTCTTATC	1440
CCTGGTGTCC	CTCCCAATTT	AGGCATACAG	CTCTTGACAT	GGGCAGTTAT	GTAGGACCCA	1500
CTCCCCACCA	CCCTTGCCAG	GGCCCCAAGT	TTGTAAATGG	CTGAGGGAAA	AGAGAGACAG	1560
AGGAGAGAGA	GAGAAATGGA	GGAGAAAGAG	AGAGAGACAG	AGAGGAGAGA	GAGACAGTGA	1620
GAGAGACAGA	AGAGAGAGAG	AGACAAAGAG	GAGAGAGAGA	GAGTCAAAGA	GAGAAAGAAA	1680
GAGAAAGAAA	TAGTAAAAAA	CAGTGTGCCC	TATTCCTTTA	AAAGCCAGGG	TAAATTTAAA	1740
ACCTGTACTT	GATAATTGAA	GGTCTTCTCT	GTGACCCTAT	AGCACTCCAA	TCCACTTTGT	1800
GGTCAGTGTA	AATAAGAGCA	TAGGCCGAAA	GCACTGAGGC	CATTGACAAC	CCGTAGCTTC	1860
CCTATCAAAA	ATCCTTAACC	CAGTAACCCG	CAGATGGACC	AAATGCATTG	AGTCGGTAGC	1920
GCAACTGCTT	TGCTAAAAGT	AGAAAAGTAA	CTTTTAGAGG	AAACCTCATT	GTGAGCACAC	1980
CTCACCTGTT	CAGAATTATT	CTAATAAAAA	AAGCAAAAAG	GTAGCTTACT	AACTCAAAAA	2040
TCTTAAAGTA	TGGGGCTATT	CTGTTAGAAA	AAGGTAATGT	AACTCCAACC	ACTGATAATT	2100
CCCTTAACCC	AGCAGATTTC	CTAACGGGAT	TTAAATCTTA	ATTACCATAC	AAAGGTCCGA	2160
CCAGACCTAG	GCGGAACTCC	CTTCAGGACA	GGACGATAGA	TGGTTCCTCC	CAGGTGATTG	2220

AGGAAAAAAA	CCACAATGGG	TATTCAGTAA	TTGATACGGG	GACTCTTG TG	GAAGCAGAGT	2280
TAGAAAAATT	GCCTAATAAC	TGGTCTCCTC	AAACGTGTGA	GCTGTTTGCA	CTCAGCCAAG	2340
CCTTAAAGTA	CTTACAGAAT	CAAAAGACTA	TCTCAATCCT	GATTCAAAAG	GTTAGCTACA	2400
CCCTCTCTGT	AATGCATTTG	CATAAGAACT	TGTTTATGGG	AATGCATCTT	GATGGGGCAG	2460
CTGGGTGTGT	ATAAAATAGG	AACCCAGCCC	AGCTCTAGGA	CTCACCCCTG	AGCGCAAAGG	2520
CAATGTTGGG	CATGCTGGTA	AAGGACCACT	AGAATCCAGC	AGCCCAGACC	CCTTTCTTTG	2580
TGGTCAAGAA	AGGCGGGAAA	AGGGGTGCAG	GACTGCTACA	TCGGTAAGCA	TAATAATCC	2640
GATAAACAGA	GGTCCATGGG	TGGTTACGCA	CCCTGGAAAG	GAATCACCC	CTGAGCACAA	2700
AGGCAATGTT	GGGCACGCTG	GTAAAGGACC	ACTAGAATCC	AGCAGCCTGG	ACCCCTTTCT	2760
TTGTGGTCAA	GAGAGGCAGG	AAAACAGGTG	CAGGACTGCA	ACATCAGTGA	GCATAACTAA	2820
TTCGATAAGC	AGAGGTCCAT	GGGTGGTGAT	GCACCCTGGA	AAGAATAAGC	ATTAGGACCA	2880
TAGAGGACAC	TCCAGGACTA	AAGCTCATCG	GAAAATGACT	AGGGTTGCTG	GCATCCCTAT	2940
GTTCTTTTTT	CAGATGGGAA	ACGTTCCCCG	CAAGACAAAA	ACGCCCCTAA	GACGTATTCT	3000
GGAGAATTGG	GACCAATTTG	ACCTCAGAC	ACTAAGAAAG	AAACGACTTA	TATTCTTCTG	3060
CAGTGCCGCC	TGGCACTCCT	GAGGGAAGTA	TAAATTATAA	CACCATCTTA	CAGCTAGACC	3120
TCTTTTGTAG	AAAAGGCAAA	TGGAGTGAAG	TGCCATAAGT	ACAAACTTTC	TTTTCATTA	3180
GAGACAACTC	ACAATTATGT	AAAAAGTGTG	ATTATGCCC	TACAGGAAGC	CTTCAGAGTC	3240
TACCTCCCTA	TCCCAGCATC	CCCGACTCCT	TCCCCAACTA	ATAAGGACCC	CCCTTCAACC	3300
CAAATGGTCC	AAAAGGAGAT	AGACAAAAGG	GTAAACAGTG	AACCAGAGAG	TGCCAATATT	3360
CCCCAATTAT	GACCCCTCCA	AGCAGTGGA	GGAAGAGAAT	TCGGCCCAGC	CAGAGTGCAT	3420
GTGCCTTTTT	CTCTCCAGAG	CTTAAAGCAA	ATAAAAACAG	ACTTAGGTAA	ATTCTCAGAT	3480
AACCCTGATG	GCTATATTGA	TGTTTTACAA	GGGTAGGAC	AATTCTTTGA	TCTGACATGG	3540
AGAGATATAA	TGTCCTGCT	AAATCAGACA	CTAACCCCAA	ATGAGAGAAG	TGCCACCATA	3600
ACTGCAGCCT	GAGAGTTTGG	CGATCTCTGG	TATCTCAGTC	AGGTCAATGA	TAGGATGACA	3660
ACAGAGGAAA	GAGAATGATT	CCCCACAGGC	CAGCAGGCAG	TTCCCAGTCT	AGACCCTCAT	3720
TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCTGCAGAC	ATTTGCTAAC	TTGTGTGCTA	3780
GAAGGACTAA	GGAAACTAG	GAAGAAGTCT	ATGAATTACT	CAATGATGTC	CACCATAACA	3840
CAGGGAAGGG	AAGAAAATCC	TACTGCCTTT	CTGGAGAGAC	TAAGGGAGGC	ATTGAGGAAG	3900
CGTGCCTCTC	TGTCACCTGA	CTCTTCTGAA	GGCCAACTAA	TCTTAAAGCG	TAAGTTTATC	3960
ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAA	CTTCAAAAGT	CTGCCGTAGG	CCCGGAGCAA	4020
AACTTAGAAA	CCCTATTGAA	CTTGGCAACC	TCGGTTTTTT	ATAATAGAGA	TCAGGAGGAG	4080
CAGGCGGAAC	AGGACAAACG	GGATTAAAAA	AAAGGCCACC	GCTTTAGTCA	TGACCCTCAG	4140
GCAAGTGGAC	TTTGGAGGCT	CTGGAAAAGG	GAAAAGCTGG	GCAAATTGAA	TGCCTAATAG	4200
GGCTTGCTTC	CAGTGCGGTC	TACAAGGACA	CTTTAAAAAA	GATTGTCCAA	GTAGAAGTAA	4260

GCCGCCCCCT	CGTCCATGCC	CCTTATTTCA	AGGGAATCAC	TGGAAGGCCC	ACTGCCCCAG	4320
GGGACAAAGG	TCCTCTGAGT	CAGAAGCCAC	TAACCAGATG	ATCCAGCAGC	AGGACTGAGG	4380
GTGCCTGGGG	CAAGCGCCAT	CCCATGCCAT	CACCCTCACA	GAGCCCTGGG	TATGCTTGAC	4440
CATTGAGGGC	CAGGAGGTTG	TCTCCTGGAC	ACTGGTGCGG	TCTTCTTAGT	CTTACTCTTC	4500
TGTCCCGGAC	AACTGTCCTC	CAGATCTGTC	ACTATCTGAG	GGGGTCCTAA	GACGGGCAGT	4560
CACTAGATAC	TTCTCCCAGC	CACTAAGTTA	TGACTGGGGA	GCTTTATTCT	TTTCACATGC	4620
TTTTCTAATT	ATGCTTGAAA	GCCCCACTAC	CTTGTTAGGG	AGAGACATTC	TAGCAAAAGC	4680
AGGGGCCATT	ATACACCTGA	ACATAGGAGA	AGGAACACCC	GTTTGTTGTC	CCCTGCTTGA	4740
GGAAGGAATT	AATCCTGAAG	TCTGGGCAAC	AGAAGGACAA	TATGGACGAG	CAAAGAATGC	4800
CCGTCTGTG	CAAGTTAAAC	TAAAGGATTC	CACCTCCTTT	CCCTACCAA	GGCAGTACCC	4860
CCTCAGACCC	AAGGCCAAC	AAGGACTCCA	AAAGATTGTT	AAGGACCTAA	AAGCCCAAGG	4920
CCTAGTAAAA	CCATGCAGTA	ACCCCTGCAG	TACTCCAATT	TTAGGAGTAC	AGAAACCCAA	4980
CAGACAGTGG	AGGTAGTGC	AAGATCTCAG	GATTATCAAT	GAGGCTGTTG	TTCCTCTATA	5040
GCCAGCTGTA	CCTAGCCCTT	ATACTCTGCT	TTCCCAAATA	CCAGAGGAAG	CAGAGTGGTT	5100
TACAGTCCTG	GACCTTCAGG	ATGCCTTCTT	CTGCATCCCT	GTACATCCTG	ACTCTCAATT	5160
CTTGTTTGCC	TTTGAAGATA	CTTCAAACCC	AACATCTCAA	CTCACCTGGA	CTATTTTACC	5220
CCAAGGGTTC	AGGGATAGTC	CCCATCTATT	TGGCCAGGCA	TTAGCCCAAG	ACTTGAGCCA	5280
ATCCTCATAC	CTGGACACTT	GTCCTTCGGT	AGGTGGATGA	TTTACTTTTG	GCCGCCCAT	5340
CAGAAACCTT	GTGCCATCAA	GCCACCCAAG	CGCTCTTCAA	TTTCTCTGCT	ACCTGTGGCT	5400
ACATGGTTTC	CAAACCAAAG	GCTCAACTCT	GCTCACAGCA	GGTTACTTAG	GGCTAAAATT	5460
ATCCAAAGGC	ACCAGGGCCC	TCAGTGAGGA	ACACATCCAG	CCTATACTGG	CTTATCCTCA	5520
TCCCAAAACC	CTAAAGCAAC	TAAGGGGATT	CCTTGGCGTA	ATAGGTTTCT	GCCGAAAATG	5580
GATTCCAGG	TATGGCGAAA	TAGCCAGGTC	ATTAAATACA	CTAATTAAGG	AAACTCAGAA	5640
AGCCAATACC	CATTTAGTAA	GATGGACAAC	TGAAGTAGAA	GTGGCTTTCC	AGGCCCTAAC	5700
CCAAGCCCCA	GTGTTAAGTT	TGCCAACAGG	GCAAGACTTT	TCTTCATATG	TCACAGAAAA	5760
AACAGGAATA	GCTCTAGGAG	TCCTTACACA	GATCCGAGGG	ATGAGCTTGC	AACCTGTGGC	5820
ATACCTGACT	AAGGAAATTG	ATGTAGTGGC	AAAGGGTTGA	CCTCATTGTT	TACGGGTAGT	5880
GGTGGCAGTA	GCAGTCTTAG	TATCTGAAGC	AGTTAAAATA	ATACAGGGAA	GAGATCTTAC	5940
TGTGTGGACA	TCTCATGATG	TGAATGGCAT	ACTCACTGCT	AAAGGAGACT	TGTGGCTGTC	6000
AGACAACTGT	TTACTTAAAT	GTCAGGCTCT	ATTACTTGAA	GGGCCAGTGC	TGCGACTGTG	6060
CACTTGTGCA	ACTCTTAACC	CAGCCACATT	TCTTCCAGAC	AATGAAGAAA	AGATAAAACA	6120
TAAGTGTCAA	CAAGTAATTT	CTCAAACCTA	TGCCACTCGA	GGGGACCTTT	TAGAGGTTCC	6180
TTTGACTGAT	CCCACCTCA	ACTTGTATAC	TGATGGAAGT	TCCTTTGTAG	AAAAAGGACT	6240
TCGAAAAGTG	GGGTATGCAG	TGGTCAGTGA	TAATGGAATA	CTTGAAAGTA	ATCCCCTCAC	6300
TCCAGGAACT	AGTGCTCAGC	TAGCAGAACT	AATAGCCCTC	ACTTGGGCAC	TAGAATTAGG	6360

AGAAGAAAAA	AGGGCAAATA	TATATACAGA	CTCTAAATAT	GCTTACCTAG	TCCTCCATGC	6420
CCATGCAGCA	ATATGGAAAG	AAAGGGAATT	CCTAACTTCT	GAGAGAACAC	CTATCAAACA	6480
TCAGGAAGCC	ATTAGGAAAT	TATTATTGGC	TGTACAGAAA	CCTAAAGAGG	TGGCAGTCTT	6540
AACTGCCGG	GGTCATCAGA	AAGGAAAGGA	AAGGGAAATA	GAAGAGAACT	GCCAAGCAGA	6600
TATTGAAGCC	AAAAGAGCTG	CAAGGCAGGA	CCCTCCATTA	GAAATGCTTA	TAAAACAACC	6660
CCTAGTATAG	GGTAATCCCC	TCCGGGAAAC	CAAGCCCCAG	TACTCAGCAG	GAGAAACAGA	6720
ATGGGGAACC	TCACGAGGAC	AGTTTTCTCC	CCTCGGGACG	GCTAGCCACT	GAAGAAGGGA	6780
AAATACTTTT	GCCTGCAACT	ATCCAATGGA	AATTACTTAA	AACCCTTCAT	CAAACCTTTC	6840
ACTTAGGCAT	CGATAGCACC	CATCAGATGG	CCAAATCATT	ATTACTGGA	CCAGGCCTTT	6900
TCAAACTAT	CAAGCAGATA	GTCAGGGCCT	GTGAAGTGTG	CCAGAGAAAT	AATCCCCTGC	6960
CTTATCGCCA	AGCTCCTTCA	GGAGAACAAA	GAACAGGCCA	TTACCCTGGA	GAAGACTGGC	7020
AACTGATTTT	ACCCACAAGC	CCAAACCTCA	GGGATTTCAG	TATCTACTAG	TCTGGGTAGA	7080
TACTTTCACG	GGTTGGGCAG	AGGCCTTCCC	CTGTAGGACA	GAAAAGGCCC	AAGAGGTAAT	7140
AAAGGCACTA	GTTTCATGAA	TAATTCCCAG	ATTTCGACTT	CCCCGAGGCT	TACAGAGTGA	7200
CAATAGCCCT	GCTTTCAGG	CCACAGTAAC	CCAGGGAGTA	TCCCAGGCGT	TAGGTATACG	7260
ATATCACTTA	CACTGCGCCT	GAAGGCCACA	GTCTCAGGG	AAGTCGAGA	AAATGAATGA	7320
AACACTCAAA	GGACATCTAA	AAAAGCAAAC	CCAGGAAACC	CACCTCACAT	GGCCTGCTCT	7380
GTTGCCTATA	GCCTTAAAAA	GAATCTGCAA	CTTCCCCCAA	AAAGCAGGAC	TTAGCCCATA	7440
CGAAATGCTG	TATGGAAGGC	CCTTCATAAC	CAATGACCTT	GTGCTTGACC	CAAGACAGCC	7500
AACTTAGTTG	CAGACATCAC	CTCCTTAGCC	AAATATCAAC	AAGTTCTTAA	AACATTACAA	7560
GGAACCTATC	CCTGAGAAGA	GGGAAAAGAA	CTATTCCACC	CTTGTGACAT	GGTATTAGTC	7620
AAGTCCCTTC	CCTCTAATTC	CCCATCCCTA	GATACATCCT	GGGAAGGACC	CTACCCAGTC	7680
ATTTTATCTA	CCCCAACTGC	GGTTAAAGTG	GCTGGAGTGG	AGTCTTGGAT	ACATCACACT	7740
TGAGTCAAAT	CCTGGATACT	GCCAAAGGAA	CCTGAAAATC	CAGGAGACAA	CGCTAGCTAT	7800
TCCTGTGAAC	CTCTAGAGGA	TTTGCGCCTG	CTCTTCAAAC	AACAACCAGG	AGGAAAGTAA	7860
CTAAATCAT	AAATCCCCAT	GGCCCTCCCT	TATCATATTT	TTCTCTTTAC	TGTTCTTTTA	7920
CCCTCTTTCA	CTCTCACTGC	ACCCCTCCA	TGCCGCTGTA	TGACCAGTAG	CTCCCTTAC	7980
CAAGAGTTTC	TATGGAGAAT	GCAGCGTCCC	GGAAATATTG	ATGCCCCATC	GTATAGGAGT	8040
CTTTCTAAGG	GAACCCCCAC	CTTCACTGCC	CACACCCATA	TGCCCCGCAA	CTGCTATCAC	8100
TCTGCCACTC	TTTGCATGCA	TGCAAATACT	CATTATTGGA	CAGGAAAAAT	GATTAATCCT	8160
AGTTGTCTCG	GAGGACTTGG	AGTCACTGTC	TGTTGGACTT	ACTTCACCCA	AACTGGTATG	8220
TCTGATGGGG	GTGGAGTTCA	AGATCAGGCA	AGAGAAAAAC	ATGTAAAAGA	AGTAATCTCC	8280
CAACTCACCC	GGGTACATGG	CACCTCTAGC	CCCTACAAAG	GACTAGATCT	CTCAAACTA	8340
CATGAAACCC	TCCGTACCCA	TACTCGCCTG	GTAAGCCTAT	TTAATACCAC	CCTCACTGGG	8400

CTCCATGAGG	TCTCGGCCCA	AAACCCTACT	AACTGTTGGA	TATGCCTCCC	CCTGAACTTC	8460
AGGCCATATG	TTTCAATCCC	TGTACCTGAA	CAATGGAACA	ACTTCAGCAC	AGAAATAAAC	8520
ACCACTTCCG	TTTTAGTAGG	ACCTCTTGTT	TCCAATCTGG	AAATAACCCA	TACCTCAAAC	8580
CTCACCTGTG	TAAAATTTAG	CAATACTACA	TACACAACCA	ACTCCCAATG	CATCAGGTGG	8640
GTAACCTCTC	CCACACAAAT	AGTCTGCCTA	CCCTCAGGAA	TATTTTTTGT	CTGTGGTACC	8700
TCAGCCTATC	GTTGTTTGAA	TGGCTCTTCA	GAATCTATGT	GCTTCCTCTC	ATTCTTAGTG	8760
CCCCCTATGA	CCATCTACAC	TGAACAAGAT	TTATACAGTT	ATGTCATATC	TAAGCCCCGC	8820
AACAAAAGAG	TACCCATTCT	TCCTTTTGTT	ATAGGAGCAG	GAGTGCTAGG	TGCACTAGGT	8880
ACTGGCATTG	GCGGTATCAC	AACCTCTACT	CAGTTCTACT	ACAAACTATC	TCAAGAACTA	8940
AATGGGGACA	TGGAACGGGT	CGCCGACTCC	CTGGTCACCT	TGCAAGATCA	ACTTAACTCC	9000
CTAGCAGCAG	TAGTCCTTCA	AAATCGAAGA	GCTTTAGACT	TGCTAACCGC	TGAAAGAGGG	9060
GGAACCTGTT	TATTTTTTAGG	GGAAGAATGC	TGTTATTATG	TTAATCAATC	CGGAATCGTC	9120
ACTGAGAAAG	TTAAAGAAAT	TCGAGATCGA	ATACAACGTA	GAGCAGAGGA	GCTTCGAAAC	9180
ACTGGACCCT	GGGGCCTCCT	CAGCCAATGG	ATGCCCTGGA	TTCTCCCCTT	CTTAGGACCT	9240
CTAGCAGCTA	TAATATTGCT	ACTCCTCTTT	GGACCCTGTA	TCTTTAACCT	CCTTGTTAAC	9300
TTTGTCTCTT	CCAGAATCGA	AGCTGTAAAA	CTACAAATGG	AGCCCAAGAT	GCAGTCCAAG	9360
ACTAAGATCT	ACCGCAGACC	CCTGGACCGG	CCTGCTAGCC	CACGATCTGA	TGTTAATGAC	9420
ATCAAAGGCA	CCCCTCCTGA	GGAAATCTCA	GCTGCACAAC	CTCTACTACG	CCCCAATTCA	9480
GCAGGAAGCA	GTTAGAGCGG	TCTCGGCCAA	CCTCCCCAAC	AGCACTTAGG	TTTTCTTGTT	9540
GAGATGGGGG	ACTGAGAGAC	AGGACTAGCT	GGATTTCCTA	GGCTGACTAA	GAATCCCTAA	9600
GCCTAGCTGG	GAAGGTGACC	ACATCCACCT	TTAAACACGG	GGCTTGCAAC	TTAGCTCACA	9660
CCTGACCAAT	CAGAGAGCTC	ACTAAAATGC	TAATTAGGCA	AAGACAGGAG	GTAAAGAAAT	9720
AGCCAATCAT	CTATTGCCTG	AGAGCACAGC	AGGAGGGACA	ATGATCGGGA	TATAAACCCA	9780
AGTCTTCGAG	CCGGCAACGG	CAACCCCCTT	TGGGTCCCCT	CCCTTTGTAT	GGGAGCTCTG	9840
TTTTCATGCT	ATTTCACTCT	ATTAAATCTT	GCAACTGCAC	TCTTCTGGTC	CATGTTTCTT	9900
ACGGCTTGAG	CTGAGCTTTC	GCTCGCCATC	CACCACTGCT	GTTTGCCGCC	ACCGCAGACC	9960
CGCCGCTGAC	TCCCATCCCT	CTGGATCATG	CAGGGTGTCC	GCTGTGCTCC	TGATCCAGCG	10020
AGGCACCCAT	TGCCGCTCCC	AATCGGGCTA	AAGGCTTGCC	ATTGTTCTTG	CATGGCTAAG	10080
TGCCTGGGTT	CATCCTAATT	GAGCTGAACA	CTAGTCACTG	GGTTCCATGG	TTCTCTTCTG	10140
TGACCCACAG	CTTCTAATAG	AGCTATAACA	CTCACCGCAT	GGCCCAAGGT	TCCATTCTTT	10200
GAATCCATAA	GGCCAAGAAC	CCCAGGTCAG	AGAACACGAG	GCTTGCCACC	ATCTTGGGAG	10260
CTCTGTGAGC	AAGGACCCCC	AAGTAACACA	ACCATGAGGG	TGCAAATGCA	TGGGCCACTA	10320
ATGGTAGAGC	AAGAAAACAG	AAGGGCCCTG	GTTCTCTGAA	GGCATCAGTG	AGCTGAAATG	10380
CCTGCCCTGG	ATGTCCTATT	CCTAGGTGTT	TTTCTGCCTG	AAGCAGATTA	AACCCCTTGT	10440
TCACCTCTCC	AAGTAGGGCT	TCTATTACAG	CCCAAATCAA	TCCCCACCCC	AGATGACAT	10499

(2) INFORMATIONS POUR LA SEQ ID NO: 4: HE2

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 2784 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:

CTCCTTCAGGAGAACAAAGAACAGGCCACTACCCAAGAGAAGACTGGCAACTAGATTTTACCCATATGCCCCAA
 ATCTCAGGGATTTTCACTATCTACTAGTTTGGGTAGATACTTTCACTGGTTGGGCAGAGGCCTTCCCCTGTAGG
 ACAGAAAAGGCCCAAGAGGTAATAAACGTTTCATGAAATAATTCCCAGATTCCGGACTTCCCCAAGGCTTACAGA
 GTGACAATGGCCCTGCTTCAAGGCTACAGTAACCCAAGGAGTATCCCAGGTGTTAGGTATACAATATCACTC
 AACTGCGCCTGGAGGCCACAGTCCTCAGGAAAGGTGGAGAAAATGAACAAAACACTCAAATGACATCTAAAA
 AAGCTAATCCAGGAAACCCACCTCGCATGGCCTGCTCTGTTGCCTATAGCCTTACTAAGAATCCGAAACTCTC
 CCAAAAAGCAGGACTTAGTCCATACAAAATGCTGTATGGACGGCCCTTCCCTAACCAATGAACTGGGCTTGA
 CCGAGAGACAGCCAACTTAGTTGCAGACATCATCTCCTTAGCCAAATATCAACAGGTTCTTAAACATTACAG
 GGAGCCTGTCCCCAAGAAGAGGGGAAAGGAACTATTCCACCCTGGTGACATGGTATTAGTCAAGTCCCTTCCCT
 CTAATTCCCCATCCCTAGATACATCCTGGGAAGGAACTACCCAGCCATTTTATCTACCTAACGGCAGTTAA
 AGTGGCTGGAGCGGAGTCTTGGATACATCACTCAAGTCAAACCCTGGATACTGCCAAAGGAACTCAAAAAT
 CCATGAGACAATGCTAGCTATTCTGTGAACCTCTAGAGGATCTGCGCCTGCTCTTCAAATGACAACCAGGGG
 GAAAGTAACTAAAAATCGTAAATCCCCTGGCCCTCCCTTATCATATTTTCTCTTTACTGTTCTCTTACCCCT
 TTCACTCTCACTGCACCCCGTCCATGCCACTGCACCCCGTCCATGCCCGTCCATGCCAGTAGCTCCCTTAG
 CAAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATTGATGCCCATTTGTATAGGAGTTTATCTAAGGGAA
 CCCCCACCTTCACTGCCCACACCCATATGCCCCACAACCTGCTATACTCTGCCACTCTTTCATGTCATGCAAA
 TACTCATTATTGGACAGGAAAAACGATTAATCCAGTTGTCTGGAGGACTTGGAGGACTCACTTCACTCATA
 CCAGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAACAGAAAAACACATAAAGGAAGTAATCTCCCACT
 GACCTGGGTACATAGCACCCCTGGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCCTCCATACC
 CATACTGGCCTGGTAAGCCTATTTAATACCACCTGACTGGGCTCCATGAGGTCTCGGCCCAAAACCCCTACTA
 ACTGTTGGATGTGCCTCCCCCTGCACCTTTAGGCCATACATTTCAATCCCTATACCTGAACAATGGAACAACCTT
 CAGCACAGAAATAAACACCACTTCTGTTTTAGTAGGTCTCTTTCCAATCTGGAAATAACCCATACCTCAAAC
 CTCACCTGTGTAAATTTAGCAATACTATAGACACAGCCAACTCCCAATGCATCAGGTGGGTAACCTCCCA
 CACGAATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTC
 TTCAGAATCTGTGTGCTTCTCTCATTCTTAGTGGCCCTATGCCATCTACACTGAACAAGATTTATACAAT
 CATGTCATACCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCTTTGTTATTGGAGCAGGAGTGCTAGGCG
 GAGTAGCTACTGGCATTGGCGGTATCAACCTCTACTCAGTTCTACTACAACTGTCTCAAGAATAAATGG
 TGACATGGAATGGGTGCTGATACCCTGGTCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT
 CAAAATCGAAGAGCTTTAGACTTGCTAACC CGGAAAGCGGGGAACTTTTTATTTTATAGAGGAAAAATGCT
 GTTGTATGTTAATCAATCCGGAATCATCACCGAGAAAGTTAAAGAAATTCAAGGTGGAATATAACGTAGAGC
 AAAGGAGCTGCAAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCTTCTTAGGA
 CCTCTAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTGTTAAGTTTGTCTTTT
 CCAGAATCGAAGCAGTAAACTACAAATCGTTCTTCAAATGGAGCCCAGATGCAGTCCATGAGTAAATCTA
 CCACGGACCCCTGGACCGGCTGCTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCGAGGAA
 ATCTCAACTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGTTGTTGGCCAACCTCCC
 CAACAGCAGTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGACAGGAATAACTAGATTTCTAGACCAACTA
 AGAATCCCTAAGACTAGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCC
 CAACCAATCAGATACTAAAGAGAGCTCACTAAATGCTAATTAGGCAAAAACAGGAGATAAAGAAATAGCCAA
 TCATCTGTTG

(2) INFORMATIONS POUR LA SEQ ID NO: 5: HE3

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 1799 paires d bases

- (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 5:

GGGATTCTTAGTCGGCCTAGGAAATCCAGCTAATCCTGTCTCTCAGTCCCCCACTCAACAGGAAACCCAAG
 TGCTGTTGGGGAGGTTGGCTGACGACCAGTCTAATGCTTCTGCGGAATTGGGGCATAGTAGGGGTTGTGCA
 GTTGAGATTTCTCGGGAGGGGTGCGTTCGATATCATTACAATTGGAGCATGGGCTAGTAGCCGGTCCAGGG
 GTCCACGGTAGATCTTAGTCATGGACTTCATCTGGGGTTCATTTGAAGAACGATTTGTAGCTTTACAACCTT
 GATTCTGGAAGAGACAACTTAACAAGGAGGTTAAAGATACAGGGTCCAAAGAGGAGTATCAATATTAGAGCT
 GCTAGAGATCCTAAGAAGGGGAGAATCCAGGGCATCCATTGGCTGAGGAGGCCCCAGGGTCTGGTGTTTTGA
 AGCTCCTCTGTTCTACGTTGTATTCAATCTCGAATTTCTTCAACTTTCTCTGTGACAATTCAGGATTGATTAA
 CATAATAACAACATTTCTTCGCTAAAATAACATAATAACAACATTTCTTCCCCTAAAAATAAACAGCTTCCCC
 TCTTTTCAGAGGTTAGCAAGTCTAAAGCTCTTCAATTTTGAAGGACTACTGATGCTAGGAAGTTAAGTTGATCT
 TGCAAGGTGACCAGGGAGTCGGCAACCCATTCCATGTCACCATTGAGTTCTTGAGATAGTTTGTAGTAGAACT
 GAGTAGAGGTTGTGGTACCGCCAATGCCAGAACCTAGTCCACCTAGCACTCCTGCTCCGATAACAAAAGGAAG
 AATGAGTACTCTTTTGTGTGGGGCTTAGGTACAACATAATTGTATAAATCTTGTTTCAGTGTAATGGTCATG
 GGGGCACTAAGAATGAGAGGAAGCACATAGATTCTGAAGAGCCATTCAAACAACGATAGGCTAAGGTACCACA
 GACAAAAAATATTCTGAGGGTAGGCAGACTATTCGTGTGGGAGGAGTTACCCACCTGATGCATTGGGAGTTG
 GTTGTGTCTACAGTATTGCTAAATTTTACACAGGTGAGGTTTGAAGTATGGGTTATTTCCAGATTGGAAACAA
 GAGGTCTACTAAACGGAAGTGGTGTATTTCTGTGCTGTAGTTGTTCCATTGTTTCAGGTACAGGGATTGA
 AATGCATGGCCTGAAATACAGGGGGAGGCACAACCAACAGTTAGTAGGGTTTTGGACCGAGACCTCATGGAGC
 CCAGTGAGGGTGGTATTAAATAGGCTTACCAGGCAAGTATGGGTATGGAGGGTTTCATGTAGTTTTAAGAGAT
 CTAGTCCTTTGTAGGGGCTAGGGGTGCTATGTACCCGGGTGAGTTGGGAGGTTACTTCCTTTACATGTTTTTC
 TCTTGCTGATCTTGAACCTCCACCCCTCAGACATACCAGTATGGGTGAAGTAAGTCCGACAGACAGTGGCT
 CCAAGTCTTCCAGGACAACCTAGGATTAATCATTTTCCCTGTCCAATAATGAGTATTGTCATGTCATGCAAAGAG
 TGGCAGAGTTATAGCAGTTGTGGGGCATATGGGTGTGGGCAGTGAAGGTGGAGTTTCTTTAGGTAACTCCT
 ATTTGATGGGGCATCAATATTTCTGGGAAGCCGATTCTTCATAGAAACTCTTGGTAAGGGGAGCTGCTGGTT
 GTACAGCAGCATGGAGGGGTGCAGTGAGAGTGAAAGGGGGTAAGAGAACAGTAAAGAGAAAAATATGATAAG
 GGAGGGCCATGGGGATTTACGATTTTAGTTACTTTCTCACGGTTGT

(2) INFORMATIONS POUR LA SEQ ID NO: 6: HG3

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 1489 paires de bases
 (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 6:

TGGTGCTTGC	CCCGGGCACT	CTCAGTCCTG	CTGCTGGATC	ATCTGGTTAG	TGGCTTCTGA	60
CTCAGAGGAC	CTACGTCCCC	TGGGGCAGTG	GGCCTTACAG	TGATTCCCTT	GACACGAGGT	120
GCATGGACGA	GGGGGCGGCT	TATTTCTATT	TGGACAATCT	TTTTTAAAGT	GTCCTTGTAG	180
ACCGCACTGG	AAGCAAACCC	TATTAGGCAT	TTGATTTGCC	TAGCTTTTCC	CTTTCCAGT	240
GCCTCCAAAG	TCCGCTTGCC	TGAGGGCCAT	GACTAAAGCG	GTGGCCTTTT	TTTTATCCCA	300
TTTGTCCCAT	TCTGCCTGCT	CATCCTGATC	TCTATTATAA	AAAAGTGGAG	TTGCCAAGTT	360
CAATAGGGTT	TCTAAGTTTT	GTTCCGGGCC	TAAGGCAGAC	TTTTGAAGTT	TTTTCCTAAT	420

GTCTGTAGCT GACTGAGTGA TAAACTTATC CTTTAAGATT AGTTGGCCTT CAGTAGAGTC	480
AGTTGACAGA GAGAGGTATG CTTCCCTCAAT GCCTCCGTTA GTCACCTCCAG AAAGGCGGTA	540
GGATTTTCTT CCTTTCCCTG TGTATAGTG GACATCATTG AATAACTCAC AGGCTTCTTT	600
CTAGTTTTCC TTAGTCCTTC TAGCACGCAA GTTAGCAAAT GTCTGCGGCA CCAATCTCCA	660
TGTTCTGATT CTGTGTCCCA GTGAGGGTCT AACTGGGAA CTGCCTGCTG GCCTGTGGGG	720
AATCGTTCTC TTTCTCTGT TGTGACCTA TCATTGACCT GACTGAGATA CCAGAGATCG	780
CCAACTCTC AGGCTGCAGT TACGGCGACA CTTCTGTCAT TTGGGGTTAG TGTCTGATTT	840
AGCAGTAACA TTATATCTCT CCATATCAGA TCAAAGGATT GTCCTAAACC TTGTAAACA	900
TCAATATAGC CATTAGGGTT ATCTGAGAAT TTACCTAGGT CTATTTTAAT TTAAAGTCTG	960
GGAGAGAAAA AGGCACATGC ACTCTGGCTG GGCCGAATTC TCTTCCTCCC ACTGCGTCTG	1020
AGAGAGAAAA AGGTACGTGC ACTCTGGCTG GGCCGAATTC TCCTCCCACC GCTTGGAGGG	1080
GGCATAATCG GGGAAATTG GCATTCTTTG GTTAGTTGTT TACCCCTTTG TCTATCTCCT	1140
TTTGGACCGT TTGGGTTGAA GGGGGTCTT TATTATTTGG GGAAGGAGTC TGGGGGATGC	1200
TGGGGTAGGG AGGTAGACTC TGAGGGCTTC CTGTAGGGCA TAAATCACAC TTTTACATA	1260
ATTGCGAGTT GTCTCTTAAT GAAAAGAAAG TTTGTACGTA TGACACTTCA CACCATTTCG	1320
CTTCTTTTCT ACAAAGAGG TCTAGCTGTA AGATGGTGTT ATAATTTATG CTTCCCTCAG	1380
GATGCCAGGT TTCTCCCCCT TAAAGAGTAT ATCGTTGCCA GGCGGTACTG CAGAAGAATA	1440
TGTCTTTTTT TTCTTAGCAT CTGAGAGTCA AATTGGTCCC AATTCTCCA	1489

(2) INFORMATIONS POUR LA SEQ ID NO: 7: **HE4**

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 1216 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 7:

TAAAGATACA GGGATTGAAA TGTATGGCCT GAAGTGCAGG GTCATATAGG TGTGGGTGGT	60
GAAAATGGGG TTTCTTTTAG AAAAATCCT ATACGATGGG TCATCAATAT TTCCAGGAAG	120
CCGCATTCTC CATAGAAGCT CTTGGAATG GGAGCTACTG GTAGTACAGT GGCATGGAGG	180
GGGTGCAGTG AGAGTGAAAG AGGGTAAAAG AACAGTAAAG AGAAAAATAT GATAAGGGAG	240
GGGTTCAAGT AGAGTGAAAG GGGTAAGAG AACAGTAAAG AAAAAATAT GACAAGGAGG	300
GCCATGAGGA TCTACGATTC TAGTTACTTT CCTCACGGTT GTCGCTTGAA GAGCAGGTGC	360
AGATCCTCTA GAGGTTTACA GGAATAGCTA GCGTTGTCTC CTGGATTTTC GGGTTCCTTT	420
GGCAGTATAC AGAGTTTGAC TCGAGTGTGA TGTATTCAAG ACTCCACTCC AGCCACTTTA	480
ACCGCAGTTG GGGTAGATAA AATGACTGGG TAGGGTCTT CCCAGGATGT ATCTAAGGAT	540
GGGGACTTAG AAGGAAGGGA CTTGACTAAT ACCATGTCAC CAGGGTGCAA TAATTACTTT	600

CCCTCTTCTC GGGAACAGGT TCCCTGTAAT GTTTTAAGAA CTTGTTGATA TTTGGCCAAG	660
GAGGTGATGT CTGCAACTAA GCTGGCCATC TCTCGGTCAA GCACAAGGTC CTTGGTTAGG	720
AAGGGCCATC CATAAGCAT TTTGTATGGG CTAAGTCTG CTTTTTGGGG AGAGTTTTGG	780
ATTCTTAGTA AGGCTGTAGG CAACAGAGCA GGCCATGCAA GGTGGGTTTC TTGGGTTAGC	840
TTTTTTAAAT GTCGTTTGAG TGCTTCATTC ATTTTCTTGA CTTTTCTGA GGATTGTGGC	900
CTCCACGCGC AGTGTAAGTG ATATTGTATG CCTAATGCCT GGGATACTCC CTGGGTTACT	960
GTAGCCTTGA AAACGGGGCC ATTGTCACTC TGTAAGCCTC GGGGAAGTCC GAATCTGGGA	1020
ATTATTTTCAT GAATTAGTGC CTTTATTACA TCTTGGTCCT TTTCTGTCCT ACAAAGGAAG	1080
GCCTCTGCCC AACCAGTGAA AATATCTACC CAGACTAGTA GATACTGAAA TCCCTGAGAT	1140
TTGGGCATGT GGGTAAAATC TAGTTGCCAG TCTTCTCCTG AGTAATGGCC TGTTCTTTGT	1200
TCTCCTGAAG GAGCTT	1216

(2) INFORMATIONS POUR LA SEQ ID NO: 8: HE5

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 976 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 8:

AGTGATAATG GAATACTTGA AAGTAATCCC CTCACTCCAG GAACTAGTGC TGAGCTGGCC	60
AAACTAATAG CCCTCACTCG GGCAC TAGAA TTAGGAGAAG AGAAAAGGGT AAATATATAT	120
ACAGACTATA AGTATGCTTA CCTAGTCCTT CATGCCCATG CAGCAATATG GAGAGAAAGG	180
GAATTCCTAA CTTCCAAAGG AACACCTATC AAACATCAGG AAGCCATTAG GATATTATTA	240
TTGGTGGTAC AGAAACCTAA AGAGGTGGCA GTCCTACACT GCTGGGGTCA TCAGAAAAAA	300
AAGGAAAGGG AAATAGAAGG GAACTACCAA GCAGATATTG AAGCCAAAAG AGCCGCAAGG	360
CAGGACCCTC CATTAGAAAT GCTTATAGAA GGACCCCTAG TGTGGGGTAA CCCCCTCCAG	420
GAAAGCAATC CCCAGTACTC AGCAGGAGAA ATAAATGGA GAACCTCAGG AGGACATACT	480
TTCTCCCTC CAGGATGGCT AGCCACCAA GAAGGAAAA TGCTTTTGCC TGCAGCTAAC	540
CAATGGAAAT TACTTAAAC CCTTCACCAA ACCTTTCACT TAGGATTGAT AGCACCCATC	600
AGATGGCCAA ATTATTATTT ACTGGATCAG GCCTTTTCAA AACTATCAAG CAGGTAGTCA	660
GGGCCTGTAA AGTGTGCCAA AGAAATAATC TCCTGCACTG CAAGCCATAC ATTTCAATCC	720
CTGTATCTTT AACCTCCTTG TTAAGTTTGT CTCTCCAGA ATCAAAGCTG TAAAACTACA	780
AATGGTTCTT CAAATGGAGT CTCAGATGCA GTCCATGACT AAGATATACC GCAGCCCCCT	840
GGAGGGGGCC TGCTAGCCCA TGCTCCAATG TTAATGACAT CGAAGGCACC CCTCCCGGGG	900
AAATCTCAAC TGCACAACCC CTACTATGTC CCAATTCAGC AGGAAGCAGT TAAAGCGGTC	960
ATCGGCCAAC CTCCTC	976

(2) INFORMATIONS POUR LA SEQ ID NO: 9: HE6

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 942 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 9:

```

AGAGGAGAAC AGCAGCATAA GCGGCTGGCA GAGGTAGGGA AAGACCAGCA AGAAGAAAAG      60
AGAGAAAGAG AAAGAGAAAG TCAGAGAAAG AGACAGAGAG AGGAAGAGAC AAAGAGACAG      120
AAAGTCAAAG AGGTAGTAGT CAGAAACAGA GACAAAAAAA AGGAGTCAGA AAGAGGGACA      180
GACACAGAAA GTCAAAAAAA AAGTTAAGAA GAAAGGAAAA GACAAAGAAG AAGTCGAAGA      240
GGAGAAAGAG AGAGATAGAA GTAGTAAAGA AAAAAACAGC ATATCCCATT CCTTTAAAGC      300
CAGGGTAAAT TTCTATCTAC CCAGCCAAGG CATATTCTAC TTATGTGGAT CTTCAACCCA      360
TATCTGCCTC TCAGACAGTT TGCAAGAAAT AATGAAATCT ATCCTTACTT TACAATCCCA      420
AATAGACTCT TTGGCAGCAG TGA CTCTCCA AACTGCAGA GGCCTAGACC TCCTCACTGC      480
TGAAAAAGGA GGACACTACA CCTTCTTAGG GGAAGAATGT TGTTTTTACA CTAACCACTC      540
GGGGATAGTA TGAGATGCTG CCCGGAGTTT ACAGGAAAAG GCTTCTGAAA TCAGACAACG      600
CCTTTCAAAT TCTTATACCA ACTTCTGGAG TTAGGCAACA TGGCTTCTCC CCTTTCTAGG      660
TCCTGTGGCA GCCATCTTGC TGTTACTCGC CTTTGGGCCC TGTATTTTTA ACCTTCTTGT      720
CAAATTTGTT TCCTCTAGAA TCGAGGCCAT CAAGCTACAG ATGGTCTTAC AAATGGAACC      780
CCAAAGAGT TCAACTAACA ACTTCTACCG AGGACCCCTG GATCAACCCA CTGGCACTTC      840
CCCTGGCCTA GAGAGTTCCC CTCTGAAGGA CACCGCAACT GCAGGGCCCT TCTTTGCCCC      900
ATCCAGCAGG AGTAGCTAGA GTGGTCATCG GCCAAATTGC CA                          942

```

(2) INFORMATIONS POUR LA SEQ ID NO: 10: HG6

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 1375 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 10:

```

CCCCAATATT CTCTTTCTGA TGGGGAAAAA TGGCCACCTG AGGGAAGCAC AAATTACAAT      60
ACTATCCTGC AGCTTGATCT TTTCTGTAAG AGGGAAGGCA AATGGAGTGA AATACCTTAT      120
GTCCAAGCTT TCTTTTCATT GAGGGAGAAT ACACAACTAT GCAAAGCTTG CAATTTACAT      180
CCCACAGGAG GACCCCTCAG CTTACCCCA TATCCTAGCC TCCCTATAGC TTCCCTTCCT      240
ATTGATGATA CTCCTCCTCT AATCTCCCCT GCCCAGAAGG AAATAAGCAA AGAAATCTCC      300
AAAGGTCCAC AAAAACCCCC GGGCTATCGG TTATGTCCCC TTCAAGCTGT AGGGGGAGGG      360

```

GAATTTGGCC CAACCCGGGT GCATGTCCCC TTCTCCCTCT CTGATTTAAA GCAGATCAGG	420
CAGACCTGGG GAAGTTTTCA GATGATCCTG ATAGGTACAT AGATGTCCTA CAGGGTCTAG	480
GGCAAACCTT TGACCTCACT TGGAGAGACG TCATGCTACT GTTAGATCAA ACCCTGGCCT	540
TTAATGAAAA GAATGCGGCT TTAGCTGCAG CCTGAGAGTT TGGAGATACC TGGTATCCTA	600
GTCAAGTAAA TGAAAGAATG ACAGCCGAAG AAAGGGACAA CTTCCCTACT GGTGAGCAAG	660
CCATCCCCAG TATGGATCCC CACTGGGACT TTGACTCAGA TCATGGGGAC TGGAGTCGTA	720
AACATCTGTT GATCTGTGTT CTGGAAGGAC TAAGGAGAAT TGGGAAAAAG CCCATGAATT	780
ATTCAATGAT ATCCACCATA ACCCAGGGAA AGGAAGAAAA TCCTTCTGCC TTCCTCGAGC	840
GGCTACAAGA GGCCTTAAGA AAATATACTC CCCTGTCACC CGAATCACTC GAGGGTCAAT	900
TGATTCTAAA AGATAAGTTT ATTACCCAAT CAGCCACAGA TATCAGGAGA AAGCTCCAAA	960
AGCAAGCCCT GAGCCCTGAA CAAAATCTAG AGACATTATT AAACCTGGCA ACCTTGGTGT	1020
TCTATAATAG GGACCAAGAG GAACAGGCCC AAAAGGAAAA GCGAGATCAG AGAAAGGCCG	1080
CAGCCTTAGT CATGGCCCTC AGACAAACAA ACCTTGGTGG TTCAGAGAGG TCAGAAAATG	1140
GAGCAGGCCA ATCACCTGGT ACGGCTTGTT ATCAGTGC GG TTTACTAGGA CACTTTAAAA	1200
AAGATTGTCC AATAAGAAAC AAGCTGCCCC CTCATCCGTG TCCACTATGC CGAGGCAATC	1260
ACTGGAAGGT GCACTGCCCC AGAGGATGAA GGTTCCTGG GTTAGAAGCC CCCAACCAGA	1320
TGATCCAACA ACAGGACTGA GGGTGCCCGG GGCAAGCACC AGCTCATGTC ATCAC	1375

(2) INFORMATIONS POUR LA SEQ ID NO: 11: HE7

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 944 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 11:

ACCTAGGAGG AACTGTCTTC AGGACAGGAC TATAGATGCT TCCTCCCAGG CGATTAAGGG	60
AAAAAGACAC AATGGGTATT CAGTAAGTGA TAAGGAAACT CTTGTAGAAG CAGAGTTAGG	120
AAAATTGCCT AATAATTGGT CTGCTCAAAT GTGCGAGCTG TTTGCACTCA GCCAAACCTT	180
AAAAGTATTA CAGAATCAGG AAGAAGCCAT CTATACCAAT TCTAAGTTAA TATGGACTGA	240
ACGAGAACTT ATTAATAGCA AAGAATAATT GAAATCCCAA ACTTACAAGG TTTTCAACAA	300
AAGCACAGTT TGCTAAAAGT TAACTGTGTA ACATGTATTA TCCTACTACC ACAAACTCTC	360
AAATGATTTC TCAGACAGTT TGCAAGAAAC AATGAAACCT ATCCTTACTC TACAATCCCA	420
AATAGACTCT TTGGCAGCAG TGACTCTCCA AAACCACCAA GGCCTAGACC TCCTCACTGC	480
TGAGAAAGGA GGA CTCTGCA CCTTCTTAGG GGAAGATTGT TGTTTTTACA CTAACCAAGTC	540
AGGGATAGTG TGAGATGCCA CCCAGCGTTT ACAGGAAAAG GCTTCTGAAA TCAGACACAA	600
TGCTTTTCAA ACCTTATAGC AACCTCTGGA GTTCGGCGAC TGGCTTTTCC CCTTTCTAGG	660

TCCTGTGACA GCCATCTTGC TATTACTCGC CTTGGGGCCC TGTATTTTAA ACCTCCTCGT	720
CAAATTTGTT TCCTCTAGGA TCGAGGCCAT CAAGCTACAG ATGGTCTTAC AAATGGAACC	780
CCAAATGAGC TCGACTAACA ACTTCTACTG AGGACCCCTG GACCGACCCA CTGGCCCTTT	840
AACTGGCTTA AAGAGTTTCC CTCTGGAGGA CACTACAAC TGCAGGGCCCC TTCTTTGCCC	900
CATCCACAGG AAGTTAGCTA GAGCAGTCAT CACCCAATTC CCAA	944

(2) INFORMATIONS POUR LA SEQ ID NO: 12: HEB

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 963 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 12:

TACAGGAACC CCATAATACG TCCTTGCCAA ATTCTATTCA GCTCCAACTG CTAGGAGTGG	60
CCCATTGTGC CTGAACCCTC AAATCATGGG AATGAGAAAT GAATTTAGAC TGACCACAGC	120
CCTTATGAGT TTTCAGCTAC AGGGGTGTAT AGAACCCCTGA TAAGGAGTTT TCTTTGTGTG	180
TGGAAGATCC TTCTATATTT GCCTCCCCAC CAACTGGACA GGAACCTGTA CTTTAGCCTA	240
CATAGTACCT CCTGTGACTT ATCCTTTTCA GAAGAGGCAG TAGCTGTGCC CATTGATGCT	300
AAGCTTCAGC CGAGAGCAAT CTCACTACTT CCTCTATTGG CTGGTTTAGG ATTTACTACC	360
ACCTAGGAAG TGGACTCACA GCCTAGATGA AATCTCTCTC CAACTTACTC AAATCCAGGA	420
CCAAATAGAC TCATTAGCAG CTGTGGTTCT CCGAACCAGT GAGCACTAGA TCTCCAATCT	480
CCTCACTGCC GAAAGGGGAG GAACATGCCT TTTTCTGAAC AAGGAATGTT GTTTTTATGT	540
CAATAAATCA GGCATAGTGA GAGATGGAAT TAAATGACTT CAGGATAGAG CTAGCAGACT	600
ACATGGTGGG ACAACCGAAA CTACCTCAGG GTTCTCACAG CCTGTTCTCC ACTGGCTTCT	660
TCCATTTTAA GGTCCCTTCC TTATGATTAT TCTAGGAGTA ACCTTTGGCC CATGTCTTTT	720
CAGTTCCTTC ATCCTTTTCGTT TTCTTCCTGA ATAGAATCAA TGAACTAGA AATGTTACTG	780
CAGATGGAAC CTCAGATGAC TTCAACCAGC ACCTATTATC AAGGACCCCT AAACCAGCCT	840
GCCGGCCCAT ACCCGGACGT TGACACCCAA ACCACCTCTC ACGAGGAAAC CTCAGCTACA	900
GAACCCCTTC TATGCCCTTA TTCAGCAGGA AGCAATTAGA GTGGTCATCC TCCACACCC	960
CAA	963

(2) INFORMATIONS POUR LA SEQ ID NO: 13: HGB

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 1362 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 13:

CCACAATATC CTCTTCAGG AGGAGAACGA TGGCCACCTG AGGGAAGTAT ACACTATAAT	60
ACCATCCTGC AACTAGATCT GTTTTGTAAG CAAGAAGGCA AGTGGATTGA GGTACCATAT	120
GTTTCAGACCT TTTTCTCATT AAGGGATGAT AACCCACGAT TGTGTAAGAC ATGTAACCTG	180
CACCCACAG GGAGTCCTCA AATTCTACCC CCATACCCAG TCCTCCCCAC GGCTCCTCCT	240
ACTAATGCCA AACCCTCTCT GGCTTCTACA GCCCAAAGG GAACAAATAA AAGAGCCTTC	300
AGAGAGCCAA GAGACCCAC TGGCCCCTGG CTATGTCTC TTCAGGCTGT AGGAGGGGAA	360
TTTGGCCCA CCCGAGTACA TGTTCCCTTT TCTCTCTCTG ATCTAAAGCA AATTAAGGCA	420
GACTTGATG AAAGTTCTCA GATGACCCCA ATAGATACGT AGATGGCCTG CTGGGTCTGG	480
GACAATCTTT TGACCTTTCC TGGAGAGAGA TCATGTTATT GCTTGATCAG ACCTAACCTC	540
TAATGAGAAG AATGCTGCTT TAACAGGAGC CCGAGAGTTT GGGGATACCT GGTACCTCAG	600
TTAAGTAAGT GATAGAATGA CATCAGAAGA GAGCAGTTT CTACTGGCCA GCAAGCAGTC	660
CCCAGTATGG ATCCCCACTG GGACCCTGAC TCGGATCATG GGGACTGGAG TCACAAACAT	720
TTACTGACCT GTATCCTAGA AGGGTTAAGG AGAACTAGGA AAAAGCCCAT GAACTATTCA	780
ATGATGTCTA CTATAACCCA AGGGAAGGAA GAAAACCTA TTGCCTTCCT CAAAAGGCTG	840
AGGGAGGCTT TGAGAAAATA TACTCCCCTG TCACCAGATT CCCTCGAAGG CCAGTTAATT	900
TTAAAGGACA AATTTATTAC TCAGTCAGCT GCAGACATTA GGAAAAAGCT CAAAAGTTA	960
GCCTTGGGCC GAGCAAAATT TGGAGGCATC ATTAAACCTG GCAACCTCAG TGTTCTATCA	1020
TAGGGACCAA GAGGAACAGG CCGAAAAGGA AAAGCAGGAT AAGAGAAAGG CTGCAGATTT	1080
AGTCATGCCC TCAGACAAAC CTTGGCGGTT CAAAGAGGAG AAAAAATGGA GCAGGCCAAT	1140
CACCCAGCAG GGCTTATTAT CAGTGCAGTT TACAAGGACA CTTTAAACAA GATTGTCCAA	1200
AGAGAAATAA GCCGCCCTCT CACCCATGTC CACTATGCCA AGGTGATCAC TGAAGGCAC	1260
ACTGTCCCAG AGGACAAAGG TTCTCTGGGC CAGAAGTCCC CAACCAGATG ATCCAGCAAC	1320
AGGATGGAGG GTGCCCCGGG CAAGCACCAG CTCGTGTTGT CA	1362

(2) INFORMATIONS POUR LA SEQ ID NO: 14: HE9

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 945 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 14:

TTGCAGATCA ATCTCAGACT GCTGTGCTAG CAATGAGTGA GGCTTCGTGG GCATGGGACC	60
CTCTGAGCCA GGCATGGGAT ATAATGTCCT TGTGTGCCAT TTGCTAAGAC TGTTGGAATA	120
GCACAGTATT AGGGTGGGAG TGGCCCGATT TTCCAGGTGC TGTCTGTCAC CGCTTCCCTT	180
GGCTAGGAAA GAGAATTCCC TGACCCCTTG TTCTTCCCAG GTAAGGCAGT GCCTCACCTT	240

GCTTCAGCTC	AACTCAGGT	GACTGCACCC	ACTGTCCTGC	CCCCACTGTC	GGACAAGCCC	300
CAGTGAGATG	AACCTGGTAC	CTCAGTTGGA	AATGCAGAAA	TCACCTGTCT	TCTGCGTCAC	360
TCACACTGGG	AGCTGTAGAC	TGGAGCTGTT	CCTATTGGC	CATCTTGAA	CCATCTCCCA	420
AATAGACTCT	TTGGCAGCAG	TGACTCTCCA	AAACCACCAA	GGCCTAGACC	TCCTCATTCG	480
TGAGAAAGGA	GGACTCTGCA	CCTTCTTAGG	GGAGGAGTGT	TGTTTTTATA	CTGACCACTC	540
AGGGATGGTA	CGAGATGCCA	CCCGATGTTT	ACAGGAAAAG	GCTTCTGAAA	TCACACAACA	600
CCTTTCAAAC	TCTTATACCA	ACCTCTGGAG	TTGGGCAACA	TGGCTTCTCC	CCTTTCTCGG	660
TCCCATTCGA	GCCATCTTGC	TATTACTCGC	CTTCAGGCTG	TGTATTTTTA	ACCTCCTTGT	720
CAAATTTGTT	TCCTCTAGAA	TTGAGGCCGT	CAAGCTACAG	ATGGTCTTAC	AAATGGGACC	780
CCAAATGAGC	TCAACTAACA	ACTTCTGCCA	AGGACCCCTG	GACCAACCTG	CTGGCCCTTT	840
CACTGGCCTT	AAGAGTTCCC	CTCTGGAGGG	CACTACAAC	GCAGGGCCCC	TTCTTTGCCC	900
CTATCCAGCA	GGAAGTAGCT	AGAGCAGTCA	TCACCCAATT	CCCAA		945

(2) INFORMATIONS POUR LA SEQ ID NO: 15: HE10

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 939 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 15:

AGAGCTACCT	TGGCAAGTAC	TCTAGGAGTA	TGGGAAAATG	AAAACAACAA	ACTCACACAC	60
CATTTTAACA	TACACAATCA	GGTCTGCCCC	CCCAGCAAGG	TATATTCTTT	GTATGTGGAA	120
CATCGACCTA	TATCTGCCTC	CCCACTAACT	AGACAGCCAC	CTGAATCTTA	GTCTTTCTAA	180
GTCCCAACAG	TAACATTGCC	CCAGGAAATC	AGACCATATC	AGTATCCCTC	AAAGCTCAAG	240
TCTGTCAAGT	GAGAGCCATA	CAACTAATAC	CCCTACTTAT	AGGGTAAGGA	ATGGCTACTG	300
CTACAGGAAC	CAGAATAGCT	AGTTTGTTTA	CTTCATTATC	CTACTACCAC	AACTCTCAA	360
ATGATTTCTC	AGACAGTTTG	CAAGAAATAA	CGAAATCTAT	CCTTACTCTA	CAATCCCAA	420
TAGACTCCTT	GGCAGCAGTG	ACCCTCCAAA	ACGGCTGAGG	CCTAGACCTC	CTCACTGCCA	480
AGAAAGGAGG	ACTCTGCATT	TTCTTAGGGG	AAGAGTGTTT	TTACACTAAC	CAGTCAGGGA	540
CAGTATGAGA	TGCCACTCGG	AGTTTACAGG	AAAAGGCTTC	TGAAGTCAGA	CAATGCCTTT	600
CAAACCTCTA	ACCAAACCTC	GGAGTTGGGC	AACATGGCTT	CTCCCTTTTC	TAGGTCCCGT	660
GACAGCCATC	TTGCTATTAT	TTGCCTTTGA	GCCCTGTATT	TTTAATCTCC	TTTTCAAATT	720
TGTTTCCTCT	GGATCGAGGC	CATCGAGCTA	CAGATGGTCT	TCACAAATGG	AACCCCAAAT	780
GAGCTCAACT	AACAACCTCT	ACTGAGGACC	CCTGGACTAA	CCTGCTGACC	CTTTCCTG	840
CCTGAAGAAT	TCCCCTCTGG	AGGACACTAC	AACTGCAGGG	CTCCTTCTTT	GCCCCATACC	900
AGCAGGAAGT	AGCTAGAGCT	GTCATTGCCT	AATTCCTAA			939

(2) INFORMATIONS POUR LA SEQ ID NO: 16: HE11

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 979 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 16:

```

AGTGATAATG GAATACTTGA AAGTAATCCC CTCACTCCCC AGGAACTAGT GCTCAGCTGG      60
CAGAACTAAT AGCCCTCACT CGGGTACTAG AATCAGGAGA AGGAAAAAGG GTAAATATAT      120
ATACAGACTC TAAGTGTGCT TACCTAGTCC TCCATGCCCA TGCAGCAATA TGGAGAGAAA      180
GGGAATTCCT AACTTCCGAG GGAACACCTA TCAAACATCA GGAAGCCATT AGGAAATTAT      240
TATTGGCTGT ACAGAAACCT AAAGAGGTGG CAGTTTACCA CTGCCGGGGT CATCAGAAAG      300
GAAAGGAAAG GGAAATACAA GGGAGCCACC AAGTTGATAT TGAAGTCAA AGAGCCACAA      360
GGCTGGACCC TCCATTAGAA ATGCTTATAG GAGGACCCCT AGTATGGGGT AATCCCCTCC      420
GGGAAGCCAA GCCCCAGTAC TCAGCAGGAG AAATAGAATA GGGAACTTCA TGAGGACATA      480
CTTCCCTCCC CTCCAGATGG CTAGCCACCA ATAAAGGAAA AATACTTTTG CCTGCAGCTA      540
ACCAATAGAA ATTACTTAAA ACCCTTCATC AAACCTTCCA CTTAGGCATT GATAGCACCC      600
ATGAGATGGC CAAATTATTA TTTACTGGAC CAGGCTTTT CAAAATATC AAGCAGATAG      660
TCAGGGCCTG TAAAGTCTGC CAAAGAAATA ATCCCTGCA CTGCAGGCCA TACATTTCAA      720
TCCCTGTATC TTTAACCTCC TTCTTAAATT TGTCTCTTCC AGAATCAAAG CTGTAAAATT      780
ACAAATAGTT CTTCAAATGG AGCCACAGAT GCAGTCCATG ACTAAGATCC ACCACAGACC      840
CCTGGACCAG CCTGCTAGCC CATGCTCCAA TGTTAATGAC ATCGAAGGCA CCCCTCCTG      900
AGGAAATCTC AACTGCACAA CCCCTACTAC GCCCAATTC AGCAGAAAGC AGTTAGAGTG      960
GTCATCAGCC AACCTCCCC                                     979

```

(2) INFORMATIONS POUR LA SEQ ID NO: 17: HG11

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 1774 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 17:

```

CATGCTGGTAAAGGACCGCTAGAATCCAGCAGCCAGGACCACTTTCTTTGTGGTCAAGAAAGGTGGGAAAACA
G
GTGCAGGACTGCTACACTGGTAAGCATAACTAATCCGATAAGCAGAGGTCCATGGGTGGTTACGCACCCCTGGA
AAGGAATAAGCATTAGGACTATAGAGGACACTCTAGGACTAATGCTCATCGGAAAATGACTAGGGGTACTGGC
ATCCCTATGTTCTTTTTCAGATGGGAAATGTTCCCCCAAGGCAGAAATGCCCTAAGATGTATTCTGGAGA
AATGGGACCAATCTGACCATCAGACACTAAGAAAGAAATGACTTATATTTCTTCTGCAGTACCACCTGGCCACA
ATATCTTCTTCAAGGGGCAGAAACCTGGCCTCCTGAGGGAAGTATAAATTATAACACCATCTTACAGCTAGAC

```

CTCTTTTGTAGAAAAGAAGGCAAATGGAGTGAAGTGCCATATGTACAACTTTCTTTTCATTAAGAGATAACT
 CCCAATTATGTAAAAAGTGTGATTTATGCCCTACAGGAAGCCCTCAGAGTCTACCTCCCGACCCAGCAAGAC
 CCCAACTCCTTCTCCAATAATAAGGACCCCTTCAACCCAAATGGTCCAAAAGGAGATAGACAAAGGGGTA
 AACAAATGAACCAAAGAGTGCCAATATTACACGATTATACTCGCTCCAAGCAGTGGGAGGAGAATTTGGCCAG
 CCAGCGTGCATGTACCTTTTTCTCTCTCAGATTTAAAGCAAATTAATAAGACCTAGGTAAATTCTCAGATAA
 CCCTGATGGCTATATTGATGTTTTACAAGGGTTAGGACAATCCTTTGATCTGACATGGAGAGATATAATGTTA
 CTGCTAAATCAGACACTAACCCCAAATGAAAAAGTGCTGCCATAACAGCAGCCTGAGAGTTTGGCGAACTCT
 GGTATCTCAGTCAGGTCAATGATAGGATGACAACAGATGAAAGAGAATGATTCCCCACAGGCCAGCAGGCAGT
 TCCCAGTGTAGACCCTCATTAGGACACAGAATCAGAACTTGGAGATTGGTGCCACAGACATTTGCTAACTTGC
 GTGCTAGAAGGACTAAGGAAAAGTAGGAAGAAGCCCATGAATTATTCAATGATGTCCCTTATAACACAGGGAA
 AGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAAGGATTGAGGAAGCATACTCCCTGTACCTGA
 CTCTATTAAAGGCCAACTAATCTTAAAGGATAAGTTTATCACTCAGTCAGCTGCAGAGATTAAGAAAAAACTT
 CAAAAGTATGCCTTAGGCCAGAGCAAACTTAGAAACCCTACTGAACTTGGCAACCTCAGTTTTTTATAATA
 GAGATCAGGAAGAGCAGGGGAATGGGACAAATGGGATAAAAAAAAAAAAAAAAAAGGTGACTGCTTTAGTCGTGG
 CCCTCAGGCAAATGGACTTTGGAGGCTCCAGAAAAGGGAAAAGCTGAGCAAATTGAATGCCTAACAGGGCTTG
 CTTCTAGTGTGGTCTACAAGGACACTTTAAAAAGATTGTCCAAGTAGAAACAAGCTGCCCCCTTGTCCATGC
 CCCTTATGTCAAGGGAATCACTGGAAGGCCCACTGCCCCAGGAGATGAAGGTCCTCTGAGTCAGAAGCCACTA
 ACCAGATAATCCAGCAGCAGGACTGAGGATGCCAGGGCAAGCGCCAGCCCATGCCATCACCTCACAGAGCC
 TTGGGTATGCTTGACCATTGA

(2) INFORMATIONS POUR LA SEQ ID NO: 18: HE12

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 (A) LONGUEUR: 938 paires de bases
 (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 18:

TGTAGGAAGA ACTCCCTTCA GGACAGGACA ATAGATGGTT CCTCCAGGT GATTAAGGAA	60
AAAAGACACA GTATTAGTA AGTGATAAGG AACTCTTGT AGAAGCAGAG TTAGAAAAAT	120
TGCCTAATAA TTGGTCTGCT CAAATGTGTG AGTTGTTTGC ACTCAGCCAA ATCTTAAAGT	180
ACTTACAGAA TCAGGAAGCA GCCATCTATA CCAATTCTAA GTTAATATGG ACTAAACGAG	240
GTTTTATTAG TAGCAAAGAA AAATTAAAT CCCAACTTA CAAGGTTTTT AACTAAAGTT	300
TGCCAAAAGT TAACAGTGTA ACATGTATTA TCCTACTATC ACACACTCTC AAAGGATTTC	360
TCAGACAGTT TGCAAGAAAT AACGTAATCT ATCCTTACTC TACAGTCCCA AATAGACTCT	420
TTGGTAGCAG TGA CTCTCCA AACTGCCGA GGTCTAGACC TCCTCAATGC TGAGAAAGGA	480
GAACTCTGCA CTTCTTAGG GGAAGAGTGC TGTTTTTACA CTAACCAAGT AGGGATAGTA	540
TGAGATACTG CCTGACGTTT ACAGGAAAAG GCTTCTGAAA TCAGACAACG CTTTCAAGC	600
TCTTATACCA ACCTCTGGAG TTGGGCAACA TGCTTCTCC CTTGCTAGG TCCTGTGGCA	660
GCCATCTTGC TATTACTTGC CTTCGGGCCC TGTATTTTAA ACCTCCTTGT CAAATTTGTT	720
TCCTCTAGGA TCAAGGCCAT CAAGCTACAG ATGGTCTTAC AAATGGAACC CCAATGAGC	780
TCAACTAACA ACTTCTACTG AGGACACCTG GACTGACCCA CTGGCCCTTT CACTGGCCTA	840
AAGAGTCCC TTCTGGAGGA CACTACAACG GCAGGGCCCC GTCTTCACCC CTATCCAGCA	900
GGAAGTAGCT AGATCAGTCA TTGCCCAATT CCCAACAG	938

(2) INFORMATIONS POUR LA SEQ ID NO: 19: HG12

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 1308 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 19:

```

GATGCTTGCC CCAGGCACCC TCAGTCCTGT TGTGGATCA TCTGGTCGGG GGCTTCTGGC   60
CCAAAGAACC TTTGTCTCTT GAGGCAGTGC ACCTTCCAGT GATTGCCTCA GCATTGTGGA  120
CATGGGCAAG GGGGCAGCTT GTTTCTCACT GGACAATCTT TTTAAGGTG TCCTTCCAAA  180
CCACACTGGT AACAAGCCCT ACCAGGTGAT TGGCCTGCTC TATTTTCTGT CCTCTCTGAA  240
CCACCAAGGT TTGTCTGTCT GAGGGTCTAT ACTAAGGCTG TGGCCTTTCT CTGATCTTGC  300
TTTTCTTTT TGGCCTGTTC CTCTTGGTAC CTATTATAGA AACTGAGGT TGCCAGGTTT  360
AACAAATGGCT CCAGATTTTG TTCAGGGCAC AGGGCTCATT TTGGAGCTTT CTCCTGATAT   420
CTGCAGCTGA TTGGGTAATA AACTTATCTT TTAGGATCAA TTGACTCTCA AGAGAGTTGG   480
GTGACAGGGG AGTATATTTT CTTGAGGCCT CCCATAGCCG CTCTAGGAAG GCAGAAGGAT   540
TTTCTTCCTT TCCCTGAGTT ATAAAAGACA TCATTGAACA ACTCATGGAC TTTTCCCAA   600
TTCTCCGTAG TCCTTCTAGA ACACAGGTCA GCAGATGTTT ACGACTCCAG TCCCCATGAT   660
CTGAGTCTAG ACACCAAGTG GGATCCATAC TGGGGATGGC CTGCTGACTG GTAGGGAATT   720
TGTCCCTTTC TTTGGCTGTC ATTCTATCAT TTAATTGACT AAGATACCAA GTATCTCCAA   780
ATTCTCAGGC TGCAGCTAAA GCTGCATTCT TTTCATTAAA GGCCAGGGTT TGATCTAATA   840
GCATGACATC TCTCCAAGTG AGGTCAAAGG TTTGCCCTAG ATCCATAGGA CATCAGAGAA   900
GGAGAAGGGG ACATACACCT GAGTTAGCCA AATTCCCCTC CCTCTACAGC TTGAAGGGGA   960
CATAAGCAAT AGCCTGGGGA TTTTGTGGT CCTTTGGAGA TTTCTTTGCT TGTTTCCTTC  1020
TGGGTGGGGG AGATTAGAGG AGGCTTATCA GTAATAGGAA GGGGAGCTAT AGGGAGGCTA  1080
GGATATGGGG GTAAGCTGAG AGGTCATCTT GTGGGATGTA AATTGCAAGC TTTGCATAGT  1140
TGTGGATTTT CTTACAATG AAAATAAAGC TTGGACATAA GGTATTTTAC TCCATTTGCC  1200
TTCCCTCTTA CAGAAAAGGT CAAGCTGCAG GATAGTACTG TAATTTATAC TTCCTTCAGG  1260
TGGCCATTTT TTCCCATCAG AGAGAGAATA CTGGGGCTGG GCCATAGT   1308

```

(2) INFORMATIONS POUR LA SEQ ID NO: 20: R1

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 711 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 20:

ACTGAGAGAC AGGACTAGCT GGATTTCCTA GGCCGACTAA GAATCCCTAA GCCTAGCTGG	60
GAAGGTGACC ACGTCCACCT TTAAACACGG GGCTTGCAAC TTAGCTCACA CCTGACCAAT	120
CAGAGAGCTC ACTAAAATGC TAATTAGGCA AAGACAGGAG GTAAAGAAAT AGCCAATCAT	180
CTATTGCCTG AGAGCACAGC AGGAGGGACA ACAATCGGGA TATAAACCCA GGCATTCGAG	240
CTGGCAACAG CAGCCCCCCT TTGGGTCCCT TCCCTTTGTA TGGGAGCTGT TTTATGCTA	300
TTTCACTCTA TTAAATCTTG CAACTGCACT CTTCTGGTCC ATGTTTCTTA CGGCTCGAGC	360
TGAGCTTTTG CTCACCGTCC ACCACTGCTG TTTGCCACCA CCGCAGACCT GCCGCTGACT	420
CCCATCCCTC TGGATCCTGC AGGGTGTCCG CTGTGCTCCT GATCCAGCGA GGCGCCATT	480
GCCGCTCCCA ATTGGGCTAA AGGCTTGCCA TTGTTCTGTC ACGGCTAAGT GCCTGGGTTT	540
GTTCTAATTG AGCTGAACAC TAGTCACTGG GTTCCATGGT TCTCTTCTGT GACCCACGGC	600
TTCTAATAGA ACTATAACAC TTACCACATG GCCCAAGATT CCATTCCTTG GAATCCGTGA	660
GGCCAAGAAC TCCAGGTCAG AGAATACGAG GCTTGCCACC ATCTTGAAG C	711

(2) INFORMATIONS POUR LA SEQ ID NO: 21: R1F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (A) LONGUEUR: 711 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 21:

ACTGAGAGAC AGGACTAGCT GGATTTCCTA GGCTGACTAA GAATCCCTAA GCCTAGCTGG	60
GAAGGTGACC ACATCCACCT TTAAACACGG GGCTTGCAAC TTAGCTCACA CCTGACCAAT	120
CAGAGAGCTC ACTAAAATGC TAATTAGGCA AAGACAGGAG GTAAAGAAAT AGCCAATCAT	180
CTATTGCCTG AGAGCACAGC AGGAGGGACA ATGATCGGGA TATAAACCCA AGTCTTCGAG	240
CCGGCAACGG CAACCCCTT TGGGTCCCCT CCCTTTGTAT GGGAGCTCTG TTTTCATGCT	300
ATTTCACTCT ATTAAATCTT GCAACTGCAC TCTTCTGGTC CATGTTTCTT ACGGCTTGAG	360
CTGAGCTTTC GCTCGCCATC CACCACTGCT GTTTGCCGCC ACCGCAGACC CGCGCTGAC	420
TCCCATCCCT CTGGATCATG CAGGGTGTCC GCTGTGCTCC TGATCCAGCG AGGCACCCAT	480
TGCCGCTCCC AATCGGGCTA AAGGCTTGCC ATTGTTCTCG CATGGCTAAG TGCCTGGGTT	540
CATCCTAATT GAGCTGAACA CTAGTCACTG GGTTCATGG TTCTTTCTG TGACCCACAG	600
CTTCTAATAG AGCTATAACA CTCACCGCAT GGCCCAAGGT TCCATTCCTT GAATCCATAA	660
GGCCAAGAAC CCCAGGTCAG AGAACACGAG GCTTGCCACC ATCTTGGGAG C	711

(2) INFORMATIONS POUR LA SEQ ID NO: 22: HERV-7q (partie codante env avec trois cadres de lecture)

- (i) CARACTERISTIQUES DE LA SEQUENCE:
- (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 22:

AAGCTCCTTCAGGAGAACAAAGAACAGGCCATTACCCTGGAGAAGACTGGCAACTGATTTTACCCACAAGCCCAA
 LysLeuLeuGlnGluAsnLysGluGlnAlaIleThrLeuGluLysThrGlyAsn...PheTyrProGlnAlaGln
 SerSerPheArgArgThrLysAsnArgProLeuProTrpArgArgLeuAlaThrAspPheThrHisLysProLys
 AlaProSerGlyGluGlnArgThrGlyHisTyrProGlyGluAspTrpGlnLeuIleLeuProThrSerProAsn

ACCTCAGGGATTTTCAGTATCTACTAGTCTGGGTAGATACTTTCACGGGTGGGCAGAGGCCTTCCCCTGTAGGAC
 ThrSerGlyIleSerValSerThrSerLeuGlyArgTyrPheHisGlyLeuGlyArgGlyLeuProLeu...Asp
 ProGlnGlyPheGlnTyrLeuLeuValTrpValAspThrPheThrGlyTrpAlaGluAlaPheProCysArgThr
 LeuArgAspPheSerIleTyr...SerGly...IleLeuSerArgValGlyGlnArgProSerProValGlyGln

AGAAAAGGCCCAAGAGGTAATAAAGGCACTAGTTCATGAAATAATCCCAGATTCCGGACTTCCCCGAGGCTTACA
 ArgLysGlyProArgGlyAsnLysGlyThrSerSer...AsnAsnSerGlnIleArgThrSerProArgLeuThr
 GluLysAlaGlnGluValIleLysAlaLeuValHisGluIleProArgPheGlyLeuProArgGlyLeuGln
 LysArgProLysArg.....ArgHis...PheMETLys...PheProAspSerAspPheProGluAlaTyrArg

GAGTGACAATAGCCCTGCTTTCCAGGCCACAGTAACCCAGGGAGTATCCCAGGCGTTAGGTATACGATATCACTT
 Glu...Gln...ProCysPheProGlyHisSerAsnProGlySerIleProGlyValArgTyrThrIleSerLeu
 SerAspAsnSerProAlaPheGlnAlaThrValThrGlnGlyValSerGlnAlaLeuGlyIleArgTyrHisLeu
 ValThrIleAlaLeuLeuSerArgProGln...ProArgGluTyrProArgArg...ValTyrAspIleThrTyr

ACACTGCGCCTGAAGGCCACAGTCCTCAGGGAAGGTCGAGAAAATGAATGAAACACTCAAAGGACATCTAAAAAA
 ThrLeuArgLeuLysAlaThrValLeuArgGluGlyArgGluAsnGlu...AsnThrGlnArgThrSerLysLys
 HisCysAla...ArgProGlnSerSerGlyLysValGluLysMETAsnGluThrLeuLysGlyHisLeuLysLys
 ThrAlaProGluGlyHisSerProGlnGlyArgSerArgLys...METLysHisSerLysAspIle...LysSer

GCAAACCCAGGAAACCCACCTCACATGGCCTGCTCTGTTGCCTATAGCCTTAAAAAGAATCTGCAACTTTCCCCA
 385 395 405 415 425 435 445
 AlaAsnProGlyAsnProProHisMETAlaCysSerValAlaTyrSerLeuLysLysAsnLeuGlnLeuSerPro
 GlnThrGlnGluThrHisLeuThrTrpProAlaLeuLeuProIleAlaLeuLysArgIleCysAsnPheProGln
 LysProArgLysProThrSerHisGlyLeuLeuCysCysLeu...Pro...LysGluSerAlaThrPheProLys

AAAAGCAGGACTTAGCCCATACGAAATGCTGTATGGAAGGCCCTTCATAACCAATGACCTTGCTGTGACCCAAG
 LysSerArgThr...ProIleArgAsnAlaValTrpLysAlaLeuHisAsnGln...ProCysAla...ProLys
 LysAlaGlyLeuSerProTyrGluMETLeuTyrGlyArgProPheIleThrAsnAspLeuValLeuAspProArg
 LysGlnAspLeuAlaHisThrLysCysCysMETGluGlyProSer...ProMETThrLeuCysLeuThrGlnAsp

ACAGCCAACTTAGTTGCAGACATCACCTCCTTAGCCAAATATCAACAAGTTCTTAAACATTACAAGGAACCTAT
 ThrAlaAsnLeuValAlaAspIleThrSerLeuAlaLysTyrGlnGlnValLeuLysThrLeuGlnGlyThrTyr
 GlnProThr...LeuGlnThrSerProPro...ProAsnIleAsnLysPheLeuLysHisTyrLysGluProIle
 SerGlnLeuSerCysArgHisHisLeuLeuSerGlnIleSerThrSerSer...AsnIleThrArgAsnLeuSer

CCCTGAGAAGAGGGAAAAGAACTATTCCACCCTTGTGACATGGTATTAGTCAAGTCCCTTCCCTCTAATTCCCCA
 Pro...GluGluGlyLysGluLeuPheHisProCysAspMETValLeuValLysSerLeuProSerAsnSerPro
 ProGluLysArgGluLysAsnTyrSerThrLeuValThrTrpTyr...SerSerProPheProLeuIleProHis
 LeuArgArgGlyLysArgThrIleProProLeu...HisGlyIleSerGlnValProSerLeu...PheProIle

TCCCTAGATACATCCTGGGAAGGACCCTACCCAGTCATTTTATCTACCCCAACTGCGGTTAAAGTGGCTGGAGTG
 SerLeuAspThrSerTrpGluGlyProTyrProValIleLeuSerThrProThrAlaValLysValAlaGlyVal
 Pro...IleHisProGlyLysAspProThrGlnSerPheTyrLeuProGlnLeuArgLeuLysTrpLeuGluTrp
 ProArgTyrIleLeuGlyArgThrLeuProSerHisPheIleTyrProAsnCysGly...SerGlyTrpSerGly

GAGTCTTGATACATCACACTTGAGTCAAATCCTGGATACTGCCAAAGGAACCTGAAAATCCAGGAGACAACGCT
 GluSerTrpIleHisHisThr...ValLysSerTrpIleLeuProLysGluProGluAsnProGlyAspAsnAla
 SerLeuGlyTyrIleThrLeuGluSerAsnProGlyTyrCysGlnArgAsnLeuLysIleGlnGluThrThrLeu
 ValLeuAspThrSerHisLeuSerGlnIleLeuAspThrAlaLysGlyThr...LysSerArgArgGlnArg...

AGCTATTCCTGTGAACCTCTAGAGGATTTGCGCCTGCTCTTCAAACAACAACCAGGAGGAAAGTAACTAAAATCA
 SerTyrSerCysGluProLeuGluAspLeuArgLeuLeuPheLysGlnGlnProGlyGlyLys...LeuLysSer
 AlaIleProValAsnLeu...ArgIleCysAlaCysSerSerAsnAsnAsnGlnGluGluSerAsn...AsnHis
 LeuPheLeu...ThrSerArgGlyPheAlaProAlaLeuGlnThrThrThrArgArgLysValThrLysIleIle

TAAATCCCCATGGCCCTCCCTTATCATATTTTCTCTTTACTGTTCTTTTACCCTCTTTCACCTCTCACTGCACCC
 ...IleProMETAlaLeuProTyrHisIlePheLeuPheThrValLeuLeuProSerPheThrLeuThrAlaPro
 LysSerProTrpProSerLeuIleIlePhePheSerLeuLeuPhePheTyrProLeuSerLeuSerLeuHisPro
 AsnProHisGlyProProLeuSerTyrPheSerLeuTyrCysSerPheThrLeuPheHisSerHisCysThrPr

CCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATT
ProProCysArgCysMETThrSerSerSerProTyrGlnGluPheLeuTrpArgMETGlnArgProGlyAsnIle
 LeuHisAlaAlaVal...ProValAlaProLeuThrLysSerPheTyrGlyGluCysSerValProGluIleLeu
 SerMETProLeuTyrAspGln...LeuProLeuProArgValSerMETGluAsnAlaAlaSerArgLysTyr...

GATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCACACCCATATGCCCCGCAACTGC
AspAlaProSerTyrArgSerLeuSerLysGlyThrProThrPheThrAlaHisThrHisMETProArgAsnCys
 METProHisArgIleGlyValPheLeuArgGluProProProSerLeuProThrProIleCysProAlaThrAla
 CysProIleVal...GluSerPhe...GlyAsnProHisLeuHisCysProHisProTyrAlaProGlnLeuLeu

TATCACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCTCT
TyrHisSerAlaThrLeuCysMETHisAlaAsnThrHisTyrTrpThrGlyLysMETIleAsnProSerCysPro
 IleThrLeuProLeuPheAlaCysMETGlnIleLeuIleIleGlyGlnGluLys...LeuIleLeuValValLeu
 SerLeuCysHisSerLeuHisAlaCysLysTyrSerLeuLeuAspArgLysAsnAsp...Ser...LeuSerTrp

GGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCAACTGGTATGTCTGATGGGGGTGGAGTTCAAGAT
GlyGlyLeuGlyValThrValCysTrpThrTyrPheThrGlnThrGlyMETSerAspGlyGlyGlyValGlnAsp
 GluAspLeuGluSerLeuSerValGlyLeuThrSerProLysLeuValCysLeuMETGlyValGluPheLysIle
 ArgThrTrpSerHisCysLeuLeuAspLeuLeuHisProAsnTrpTyrVal...TrpGlyTrpSerSerArgSer

CAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCCTACAAA
GlnAlaArgGluLysHisValLysGluValIleSerGlnLeuThrArgValHisGlyThrSerSerProTyrLys
 ArgGlnGluLysAsnMET...LysLys...SerProAsnSerProGlyTyrMETAlaProLeuAlaProThrLys
 GlyLysArgLysThrCysLysArgSerAsnLeuProThrHisProGlyThrTrpHisLeu...ProLeuGlnArg

GGACTAGATCTCTCAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCTC
GlyLeuAspLeuSerLysLeuHisGluThrLeuArgThrHisThrArgLeuValSerLeuPheAsnThrThrLeu
 Asp...IleSerGlnAsnTyrMETLysProSerValProIleLeuAlaTrp...AlaTyrLeuIleProProSer
 ThrArgSerLeuLysThrThr...AsnProProTyrProTyrSerProGlyLysProIle...TyrHisProHis

ACTGGGCTCCATGAGGTCTCGGCCCAAACCCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGGCCATAT
ThrGlyLeuHisGluValSerAlaGlnAsnProThrAsnCysTrpIleCysLeuProLeuAsnPheArgProTyr
 LeuGlySerMETArgSerArgProLysThrLeuLeuThrValGlyTyrAlaSerPro...ThrSerGlyHisMET
 TrpAlaPro...GlyLeuGlyProLysProTyr...LeuLeuAspMETProProProGluLeuGlnAlaIleCys

GTTTCAATCCCTGTACCTGAACAATGGAACAACCTTCAGCACAGAAATAAACACCACTTCCGTTTGTAGTAGGACCT
ValSerIleProValProGluGlnTrpAsnAsnPheSerThrGluIleAsnThrThrSerValLeuValGlyPro
 PheGlnSerLeuTyrLeuAsnAsnGlyThrThrSerAlaGlnLys...ThrProLeuProPhe...AspLeu
 PheAsnProCysThr...ThrMETGluGlnLeuGlnHisArgAsnLysHisHisPheArgPheSerArgThrSer

CTTGTTTCCAATCTGGAAATAACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACC
LeuValSerAsnLeuGluIleThrHisThrSerAsnLeuThrCysValLysPheSerAsnThrThrTyrThrThr
 LeuPheProIleTrpLys...ProIleProGlnThrSerProVal...AsnLeuAlaIleLeuHisThrGlnPro
 CysPheGlnSerGlyAsnAsnProTyrLeuLysProHisLeuCysLysIle...GlnTyrTyrIleHisAsnGln

AACTCCCAATGCATCAGGTGGGTAACTCCTCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTGTCTGT
AsnSerGlnCysIleArgTrpValThrProThrGlnIleValCysLeuProSerGlyIlePhePheValCys
 ThrProAsnAlaSerGlyGly...LeuLeuProHisLys...SerAlaTyrProGlnGluTyrPheLeuSerVal
 LeuPromETHisGlnValGlyAsnSerSerHisThrAsnSerLeuProThrLeuArgAsnIlePheCysLeuTrp

GGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCATTCTTAGTGCCCCCTATG
GlyThrSerAlaTyrArgCysLeuAsnGlySerSerGluSerMETCysPheLeuSerPheLeuValProPromET
 ValProGlnProIleValVal...METAlaLeuGlnAsnLeuCysAlaSerSerHisSer...CysProLeu...
 TyrLeuSerLeuSerLeuPheGluTrpLeuPheArgIleTyrValLeuProLeuIleLeuSerAlaProTyrAsp

ACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCTTCCT
ThrIleTyrThrGluGlnAspLeuTyrSerTyrValIleSerLysProArgAsnLysArgValProIleLeuPro
 ProSerThrLeuAsnLysIleTyrThrValMETSerTyrLeuSerProAlaThrLysGluTyrProphePheLeu
 HisLeuHis...ThrArgPheIleGlnLeuCysHisIle...AlaProGlnGlnLysSerThrHisSerSerPhe

TTTGTTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTAC
PheValIleGlyAlaGlyValLeuGlyAlaLeuGlyThrGlyIleGlyGlyIleThrThrSerThrGlnPheTyr
 LeuLeu...GluGlnGluCys...ValHis...ValLeuAlaLeuAlaValSerGlnProLeuLeuSerSerThr
 CysTyrArgSerArgSerAlaArgCysThrArgTyrTrpHisTrpArgTyrHisAsnLeuTyrSerValLeuLeu

TACAAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTCGCCGACTCCCTGGTCACCTTGCAAGATCAACTT
TyrLysLeuSerGlnGluLeuAsnGlyAspMETGluArgValAlaAspSerLeuValThrLeuGlnAspGlnLeu
 ThrAsnTyrLeuLysAsn...METGlyThrTrpAsnGlySerProThrProTrpSerProCysLysIleAsnLeu
 GlnThrIleSerArgThrLysTrpGlyHisGlyThrGlyArgArgLeuProGlyHisLeuAlaArgSerThr...

AACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGT
AsnSerLeuAlaAlaValValLeuGlnAsnArgArgAlaLeuAspLeuLeuThrAlaGluArgGlyGlyThrCys
 ThrPro...GlnGln...SerPheLysIleGluGluLeu...ThrCys...ProLeuLysGluGlyGluProVal
 LeuProSerSerSerSerProSerLysSerLysSerPheArgLeuAlaAsnArg...LysArgGlyAsnLeuPhe

TTATTTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGA
LeuPheLeuGlyGluGluCysTyrTyrValAsnGlnSerGlyIleValThrGluLysValLysGluIleArg
 TyrPhe...GlyLysAsnAlaValIleMETLeuIleAsnProGluSerSerLeuArgLysLeuLysLysPheGlu
 IlePheArgGlyArgMETLeuLeuLeuCys...SerIleArgAsnArgHis...GluSer...ArgAsnSerArg

GATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCTGG
AspArgIleGlnArgArgAlaGluGluLeuArgAsnThrGlyProTrpGlyLeuLeuSerGlnTrpMETProTrp
 IleGluTyrAsnValGluGlnArgSerPheGluThrLeuAspProGlyAlaSerSerAlaAsnGlyCysProGly
 SerAsnThrThr...SerArgGlyAlaSerLysHisTrpThrLeuGlyProProGlnPromETAspAlaLeuAsp

ATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAACCTCCTT
IleLeuProPheLeuGlyProLeuAlaAlaIleIleLeuLeuLeuLeuPheGlyProCysIlePheAsnLeuLeu
 PheSerProSer...AspLeu...GlnLeu...TyrCysTyrSerSerLeuAspProValSerLeuThrSerLeu
 SerProLeuLeuArgThrSerSerSerTyrAsnIleAlaThrProLeuTrpThrLeuTyrLeu...ProProCys

GTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATC
ValAsnPheValSerSerArgIleGluAlaValLysLeuGlnMETGluProLysMETGlnSerLysThrLysIle
 LeuThrLeuSerLeuProGluSerLysLeu...AsnTyrLysTrpSerProArgCysSerProArgLeuArgSer
 ...LeuCysLeuPheGlnAsnArgSerCysLysThrThrAsnGlyAlaGlnAspAlaValGlnAsp...AspLeu

TACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAA
TyrArgArgProLeuAspArgProAlaSerProArgSerAspValAsnAspIleLysGlyThrProProGluGlu
 ThrAlaAspProTrpThrGlyLeuLeuAlaHisAspLeuMETLeuMETThrSerLysAlaProLeuLeuArgLys
 ProGlnThrProGlyProAlaCys...ProThrIle...Cys.....HisGlnArgHisProSer...GlyAsn

ATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCCA
IleSerAlaAlaGlnProLeuLeuArgProAsnSerAlaGlySerSer...SerGlyArgArgProThrSerPro
 SerGlnLeuHisAsnLeuTyrTyrAlaProIleGlnGlnGluAlaValArgAlaValValGlyGlnProProGln
 LeuSerCysThrThrSerThrThrProGlnPheSerArgLysGlnLeuGluArgSerSerAlaAsnLeuProAsn

ACAGCACTTAGGTTTTCTGTTGAGATGGGGG
 ThrAlaLeuArgPheSerCys...AspGlyGly
 GlnHisLeuGlyPheProValGluMETGly
 SerThr...ValPheLeuLeuArgTrpGly

(2) INFORMATIONS POUR LA SEQ ID NO: 23: HERV-7q (protéine env déduite)

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 23:

P K T A N L V A D I T S L A K Y Q Q V L K T L Q G
 CCCAAGACAGCCAACTTAGTTGCAGACATCACCTCCTTAGCCAAATATCAACAAGTTCTTAAACATTACAAGGA
 T Y P X E E G K E L F H P C D M V L V K S L P S N
 ACCTATCCCTGAGAAGAGGGAAAAGAACTATTCACCCTTGTGACATGGTATTAGTCAAGTCCCTTCCCTCTAAT
 S P S L D T S W E G P Y P V I L S T P T A V K V A
 TCCCCATCCCTAGATACATCCTGGGAAGGACCCTACCCAGTCATTTTATCTACCCCAACTGCGGTTAAAGTGGCT
 G V E S W I H H T X V K S W I L P K E P E N P G D
 GGAGTGGAGTCTTGGATACATCACACTTGAGTCAAATCCTGGATACTGCCAAAGGAACCTGAAAATCCAGGAGAC
 N A S Y S C E P L E D L R L L F K Q Q P G G K * L
 AACGCTAGCTATTCTGTGAACCTCTAGAGGATTTGCGCCTGCTCTTCAAACAACAACCAGGAGGAAAGTAAC
 K S X I P M A L P Y H I F L F T V L L P S F T L T
 AAATCATAAATCCCCATGGCCCTCCCTTATCATATTTTTCTCTTTACTGTTCTTTTACCCTCTTTCACTCTCACT
 A P P P C R C M T S S S P Y Q E F L W R M Q R P G
 GCACCCCTCCATGCCGCTGTATGACCAGTAGCTCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGTCCCGGA
 N I D A P S Y R S L S K G T P T F T A H T H M P R
 AATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCCACCTTCACTGCCACACCCATATGCCCCG
 N C Y H S A T L C M H A N T H Y W T G K M I N P S
 AACTGCTATCACTCTGCCACTCTTTGCATGCATGCAAATACTCATTATTGGACAGGAAAATGATTAATCCTAGT

C P G G L G V T V C W T Y F T Q T G M S D G G G V
 TGTCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCAACTGGTATGTCTGATGGGGGTGGAGTT
 Q D Q A R E K H V K E V I S Q L T R V H G T S S P
 CAAGATCAGGCAAGAGAAAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCC
 Y K G L D L S K L H E T L R T H T R L V S L F N T
 TACAAAGGACTAGATCTCTCAAACTACATGAAACCTCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACC
 T L T G L H E V S A Q N P T N C W I C L P L N F R
 ACCCTCACTGGGCTCCATGAGGTCTCGGCCCAAAACCTACTAACTGTTGGATATGCCTCCCCCTGAACTTCAGG
 P Y V S I P V P E Q W N N F S T E I N T T S V L V
 CCATATGTTTCAATCCCTGTACCTGAACAATGGAACAACCTTCAGCACAGAAATAAACACCACTTCCGTTTGTAGTA
 G P L V S N L E I T H T S N L T C V K F S N T T Y
 GGACCTCTTGTTCATCTGGAAATAACCCATACCTCACCTGTGTAAAATTTAGCAATACTACATAC
 T T N S Q C I R W V T P P T Q I V C L P S G I F F
 ACAACCAACTCCCATGTCATCAGGTGGGTAACCTCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTT
 V C G T T S A Y R C L N G S S E S M C F L S F L V P
 GTCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCTCTATTCTTAGTGCCC
 P M T I Y T E Q D L Y S Y V I S K P R N K R V P I
 CCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTATATCTAAGCCCCGCAACAAAAGAGTACCCATT
 L P F V I G A G V L G A L G T G I G G I T T S T Q
 CTTCTTTTGTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAG
 F Y Y K L S Q E L N G D M E R V A D S L V T L Q D
 TTCTACTACAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTGCGCGACTCCCTGGTCACCTTGCAAGAT
 Q L N S L A A V V L Q N R R A L D L L T A E R G G
 CAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGA
 T C L F L G E E C C Y Y V N Q S G I V T E K V K E
 ACCTGTTTATTTTATAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA
 I R D R I Q R R A E E L R N T G P W G L L S Q W M
 ATTCGAGATCGAATACAACGTAGAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCAATGGATG
 P W I L P F L G P L A A I I L L L L F G P C I F N
 CCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTAAC
 L L V N F V S S R I E A V K L Q M E P K M Q S K T
 CTCCTTGTTAATCTTCTCTTCCAGAATCGAAGCTGTAAAACCTACAAATGGAGCCCAAGATGCAGTCCAAGACT
 K I Y R R P L D R P A S P R S D V N D I K G T P P
 AAGATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCT
 E E I S A A Q P L L R P N S A G S S X S G R R P T
 GAGGAAATCTCAGCTGCACAACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCTGTCGGCCAACC
 S P T A L R F S C X
 TCCCAACAGCACTTAGGTTTTCTGTTGA

(2) INFORMATIONS POUR LA SEQ ID NO: 24: HERV-7q (partie codante gag)

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 24:

T S F V E K A N G V K C H K Y
 ACC TCT TTT GTA GAA AAG GCA AAT GGA GTG AAG TGC CAT AAG TAC
 K L S F H X E T T H N Y V K S
 AAA CTT TCT TTT CAT TAA GAG ACA ACT CAC AAT TAT GTA AAA AGT
 V I Y A L Q E A F R V Y L P I
 GTG ATT TAT GCC CTA CAG GAA GCC TTC AGA GTC TAC CTC CCT ATC

```

P   A   S   P   T   P   S   P   T   N   K   D   P   P   S
CCA GCA TCC CCG ACT CCT TCC CCA ACT AAT AAG GAC CCC CCT TCA
T   Q   M   V   Q   K   E   I   D   K   R   V   N   S   E
ACC CAA ATG GTC CAA AAG GAG ATA GAC AAA AGG GTA AAC AGT GAA
P   K   S   A   N   I   P   Q   L   X   P   L   Q   A   V
CCA AAG AGT GCC AAT ATT CCC CAA TTA TGA CCC CTC CAA GCA GTG
G   G   R   E   F   G   P   A   R   V   H   V   P   F   S
GGA GGA AGA GAA TTC GGC CCA GCC AGA GTG CAT GTG CCT TTT TCT
L   P   D   L   K   Q   I   K   T   D   L   G   K   F   S
CTC CCA GAC TTA AAG CAA ATA AAA ACA GAC TTA GGT AAA TTC TCA
D   N   P   D   G   Y   I   D   V   L   Q   G   L   G   Q
GAT AAC CCT GAT GGC TAT ATT GAT GTT TTA CAA GGG TTA GGA CAA
F   F   D   L   T   W   R   D   I   M   S   L   L   N   Q
TTC TTT GAT CTG ACA TGG AGA GAT ATA ATG TCA CTG CTA AAT CAG
T   L   T   P   N   E   R   S   A   T   I   T   A   A   X
ACA CTA ACC CCA AAT GAG AGA AGT GCC ACC ATA ACT GCA GCC TGA
E   F   G   D   L   W   Y   L   S   Q   V   N   D   R   M
GAG TTT GGC GAT CTC TGG TAT CTC AGT CAG GTC AAT GAT AGG ATG
T   T   E   E   R   E   X   F   P   T   G   Q   Q   A   V
ACA ACA GAG GAA AGA GAA TGA TTC CCC ACA GGC CAG CAG GCA GTT
P   S   L   D   P   H   W   D   T   E   S   E   H   G   D
CCC AGT CTA GAC CCT CAT TGG GAC ACA GAA TCA GAA CAT GGA GAT
W   C   C   R   H   L   L   T   C   V   L   E   G   L   R
TGG TGC TGC AGA CAT TTG CTA ACT TGT GTG CTA GAA GGA CTA AGG
K   T   R   K   K   S   M   N   Y   S   M   M   S   T   I
AAA ACT AGG AAG AAG TCT ATG AAT TAC TCA ATG ATG TCC ACC ATA
T   Q   G   R   E   E   N   P   T   A   F   L   E   R   L
ACA CAG GGA AGG GAA GAA AAT CCT ACT GCC TTT CTG GAG AGA CTA
R   E   A   L   R   K   R   A   S   L   S   P   D   S   S
AGG GAG GCA TTG AGG AAG CGT GCC TCT CTG TCA CCT GAC TCT TCT
E   G   Q   L   I   L   K   R   K   F   I   T   Q   S   A
GAA GGC CAA CTA ATC TTA AAG CGT AAG TTT ATC ACT CAG TCA GCT
A   D   I   R   K   K   L   Q   K   S   A   V   G   P   E
GCA GAC ATT AGA AAA AAA CTT CAA AAG TCT GCC GTA GGC CCG GAG
Q   N   L   E   T   L   L   N   L   A   T   S   V   F   Y
CAA AAC TTA GAA ACC CTA TTG AAC TTG GCA ACC TCG GTT TTT TAT
N   R   D   Q   E   E   Q   A   E   Q   D   K   R   D   X
AAT AGA GAT CAG GAG GAG CAG GCG GAA CAG GAC AAA CGG GAT TAA
K   K   G   H   R   F   S   H   D   P   Q   A   S   G   L
AAA AAA GGC CAC CGC TTT AGT CAT GAC CCT CAG GCA AGT GGA CTT
W   R   L   W   K   R   E   K   L   G   K   L   N   A   X
TGG AGG CTC TGG AAA AGG GAA AAG CTG GGC AAA TTG AAT GCC TAA

```

(2) INFORMATIONS POUR LA SEQ ID NO: 25: protéine env (cadre de lecture 1)

(i) CARACTERISTIQUES DE LA SEQUENCE:

(B) TYPE: acide aminé,

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 25:

```

PKTANLVADITSLAKYQQVLKTLQGTYPXEEGKELFHPCDMVLVKSLPSNSPSLDTSWEG
PYPVILSTPTAVKVAGVESWIHHTXVKSUILPKPENPGDNASYSCEPLEDLRLLLFKQQP
GGKXLKSXIPMALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFWRMQRPGNIDAP
SYRSLSKGTPFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWTYFT

```


QTGMSDGGGVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLETLRTHTRLVSLFNT
 TLTLGLHEVSAQNPTNCWICLPLNFRPYVSIPVPEQWNNFSTEINTTSVLVGPLVSNLEIT
 HTSNLTCVKFSNTTTYTTNSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSSESMCFL
 SFLVPPMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKYL
 SQELNGDMERVADSLVTLQDQLNSLAAVVLQNRALDLLTAERGGTCLFLGEECCYYVNO
 SGIVTEKVKEIRDRIQRRAEELRNTGPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFN
 LLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLL
 RPNSAGSSXSGRRPTSPTALRFSCX

(2) INFORMATIONS POUR LA SEQ ID NO: 26: protéine gag

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (B) TYPE: acide aminé,
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 26:

TSFVEKANGVKCHKYKLSFHKETTHNYVKSVIYALQEAFRVYLPPIASPTSPPTNKDPPS
 TOMVQKEIDKRVNSEPKSANIPQLXPLQAVGGREFGPARVHVPFSLPDLKQIKTDLGKFS
 DNPDGYIDVLQGLGQFFDLTWRDIMSLLNQTLPNERSATITAAXEFGDLWYLSQVNDRM
 TTEEREXFPTGQQA VPSLDPHWDTESEHGDWCCRHLITCVLEGLRKRTRKSMNYSMSTI
 TQGREENPTAFLERLREALRKRA SLSPDSSEGQLILKRKFITQSAADIRKKLQKSAVGPE
 QNLETLNLATS VFYNRDQEEQAEQDKRDXKKGHRFSDHPQASGLWRLWKREKLGLKNAX

(2) INFORMATIONS POUR LA SEQ ID NO: 27: protéine env (cadre de lecture 1)

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (B) TYPE: acide aminé,
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 27:

LysLeuLeuGlnGluAsnLysGluGlnAlaIleThrLeuGluLysThrGlyAsn...PheTyrProGlnAlaGln
 ThrSerGlyIleSerValSerThrSerLeuGlyArgTyrPheHisGlyLeuGlyArgGlyLeuProLeu...Asp
 ArgLysGlyProArgGlyAsnLysGlyThrSerSer...AsnAsnSerGlnIleArgThrSerProArgLeuThr
 Glu...Gln...ProCysPheProGlyHisSerAsnProGlySerIleProGlyValArgTyrThrIleSerLeu
 ThrLeuArgLeuLysAlaThrValLeuArgGluGlyArgGluAsnGlu...AsnThrGlnArgThrSerLysLys
 AlaAsnProGlyAsnProProHisMETAlaCysSerValAlaTyrSerLeuLysLysAsnLeuGlnLeuSerPro
 LysSerArgThr...ProIleArgAsnAlaValTrpLysAlaLeuHisAsnGln...ProCysAla...ProLys
 ThrAlaAsnLeuValAlaAspIleThrSerLeuAlaLysTyrGlnGlnValLeuLysThrLeuGlnGlyThrTyr
 Pro...GluGluGlyLysGluLeuPheHisProCysAspMETValLeuValLysSerLeuProSerAsnSerPro
 SerLeuAspThrSerTrpGluGlyProTyrProValIleLeuSerThrProThrAlaValLysValAlaGlyVal

GluSerTrpIleHisHisThr...ValLysSerTrpIleLeuProLysGluProGluAsnProGlyAspAsnAla
 SerTyrSerCysGluProLeuGluAspLeuArgLeuLeuPheLysGlnGlnProGlyGlyLys...LeuLysSer
 ...IleProMETAlaLeuProTyrHisIlePheLeuPheThrValLeuLeuProSerPheThrLeuThrAlaPro
ProProCysArgCysMETThrSerSerSerProTyrGlnGluPheLeuTrpArgMETGlnArgProGlyAsnIle
AspAlaProSerTyrArgSerLeuSerLysGlyThrProThrPheThrAlaHisThrHisMETProArgAsnCys
TyrHisSerAlaThrLeuCysMETHisAlaAsnThrHisTyrTrpThrGlyLysMETIleAsnProSerCysPro
GlyGlyLeuGlyValThrValCysTrpThrTyrPheThrGlnThrGlyMETSerAspGlyGlyGlyValGlnAsp
GlnAlaArgGluLysHisValLysGluValIleSerGlnLeuThrArgValHisGlyThrSerSerProTyrLys
GlyLeuAspLeuSerLysLeuHisGluThrLeuArgThrHisThrArgLeuValSerLeuPheAsnThrThrLeu
ThrGlyLeuHisGluValSerAlaGlnAsnProThrAsnCysTrpIleCysLeuProLeuAsnPheArgProTyr
ValSerIleProValProGluGlnTrpAsnAsnPheSerThrGluIleAsnThrThrSerValLeuValGlyPro
LeuValSerAsnLeuGluIleThrHisThrSerAsnLeuThrCysValLysPheSerAsnThrThrTyrThrThr
AsnSerGlnCysIleArgTrpValThrProProThrGlnIleValCysLeuProSerGlyIlePhePheValCys
GlyThrSerAlaTyrArgCysLeuAsnGlySerSerGluSerMETCysPheLeuSerPheLeuValProProMET
ThrIleTyrThrGluGlnAspLeuTyrSerTyrValIleSerLysProArgAsnLysArgValProIleLeuPro
PheValIleGlyAlaGlyValLeuGlyAlaLeuGlyThrGlyIleGlyGlyIleThrThrSerThrGlnPheTyr
TyrLysLeuSerGlnGluLeuAsnGlyAspMETGluArgValAlaAspSerLeuValThrLeuGlnAspGlnLeu
AsnSerLeuAlaAlaValValLeuGlnAsnArgArgAlaLeuAspLeuLeuThrAlaGluArgGlyGlyThrCys
LeuPheLeuGlyGluGluCysCysTyrTyrValAsnGlnSerGlyIleValThrGluLysValLysGluIleArg
AspArgIleGlnArgArgAlaGluGluLeuArgAsnThrGlyProTrpGlyLeuLeuSerGlnTrpMETProTrp
IleLeuProPheLeuGlyProLeuAlaAlaIleIleLeuLeuLeuLeuPheGlyProCysIlePheAsnLeuLeu
ValAsnPheValSerSerArgIleGluAlaValLysLeuGlnMETGluProLysMETGlnSerLysThrLysIle
TyrArgArgProLeuAspArgProAlaSerProArgSerAspValAsnAspIleLysGlyThrProProGluGlu
IleSerAlaAlaGlnProLeuLeuArgProAsnSerAlaGlySerSer...SerGlyArgArgProThrSerPro
 ThrAlaLeuArgPheSerCys...AspGlyGly

(2) INFORMATIONS POUR LA SEQ ID NO: 28: protéine env (cadre de lecture 2)

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (B) TYPE: acide aminé,
 - (D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: protéine
 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 28:

SerSerPheArgArgThrLysAsnArgProLeuProTrpArgArgLeuAlaThrAspPheThrHisLysProLys
 ProGlnGlyPheGlnTyrLeuLeuValTrpValAspThrPheThrGlyTrpAlaGluAlaPheProCysArgThr
 GluLysAlaGlnGluValIleLysAlaLeuValHisGluIleIleProArgPheGlyLeuProArgGlyLeuGln
 SerAspAsnSerProAlaPheGlnAlaThrValThrGlnGlyValSerGlnAlaLeuGlyIleArgTyrHisLeu
 HisCysAla...ArgProGlnSerSerGlyLysValGluLysMETAsnGluThrLeuLysGlyHisLeuLysLys
 GlnThrGlnGluThrHisLeuThrTrpProAlaLeuLeuProIleAlaLeuLysArgIleCysAsnPheProGln
 LysAlaGlyLeuSerProTyrGluMETLeuTyrGlyArgProPheIleThrAsnAspLeuValLeuAspProArg
 GlnProThr...LeuGlnThrSerProPro...ProAsnIleAsnLysPheLeuLysHisTyrLysGluProIle
 ProGluLysArgGluLysAsnTyrSerThrLeuValThrTrpTyr...SerSerProPheProLeuIleProHis
 Pro...IleHisProGlyLysAspProThrGlnSerPheTyrLeuProGlnLeuArgLeuLysTrpLeuGluTrp
 SerLeuGlyTyrIleThrLeuGluSerAsnProGlyTyrCysGlnArgAsnLeuLysIleGlnGluThrThrLeu
 AlaIleProValAsnLeu...ArgIleCysAlaCysSerSerAsnAsnAsnGlnGluGluSerAsn...AsnHis
 LysSerProTrpProSerLeuIleIlePhePheSerLeuLeuPhePheTyrProLeuSerLeuSerLeuHisPro
 LeuHisAlaAlaVal...ProValAlaProLeuThrLysSerPheTyrGlyGluCysSerValProGluIleLeu
 METProHisArgIleGlyValPheLeuArgGluProProProSerLeuProThrProIleCysProAlaThrAla
 IleThrLeuProLeuPheAlaCysMETGlnIleLeuIleIleGlyGlnGluLys...LeuIleLeuValValLeu
 GluAspLeuGluSerLeuSerValGlyLeuThrSerProLysLeuValCysLeuMETGlyValGluPheLysIle
 ArgGlnGluLysAsnMET...LysLys...SerProAsnSerProGlyTyrMETAlaProLeuAlaProThrLys
 Asp...IleSerGlnAsnTyrMETLysProSerValProIleLeuAlaTrp...AlaTyrLeuIleProProSer
 LeuGlySerMETArgSerArgProLysThrLeuLeuThrValGlyTyrAlaSerPro...ThrSerGlyHisMET
 PheGlnSerLeuTyrLeuAsnAsnGlyThrThrSerAlaGlnLys...ThrProLeuProPhe.....AspLeu
 LeuPheProIleTrpLys...ProIleProGlnThrSerProVal...AsnLeuAlaIleLeuHisThrGlnPro
 ThrProAsnAlaSerGlyGly...LeuLeuProHisLys...SerAlaTyrProGlnGluTyrPheLeuSerVal

ValProGlnProIleValVal...METAlaLeuGlnAsnLeuCysAlaSerSerHisSer...CysProLeu...
 ProSerThrLeuAsnLysIleTyrThrValMETSerTyrLeuSerProAlaThrLysGluTyrProPhePheLeu
 LeuLeu...GluGlnGluCys...ValHis...ValLeuAlaLeuAlaValSerGlnProLeuLeuSerSerThr
 ThrAsnTyrLeuLysAsn...METGlyThrTrpAsnGlySerProThrProTrpSerProCysLysIleAsnLeu
 ThrPro...GlnGln...SerPheLysIleGluGluLeu...ThrCys...ProLeuLysGluGlyGluProVal
 TyrPhe...GlyLysAsnAlaValIleMETLeuIleAsnProGluSerSerLeuArgLysLeuLysLysPheGlu
 IleGluTyrAsnValGluGlnArgSerPheGluThrLeuAspProGlyAlaSerSerAlaAsnGlyCysProGly
 PheSerProSer...AspLeu...GlnLeu...TyrCysTyrSerSerLeuAspProValSerLeuThrSerLeu
 LeuThrLeuSerLeuProGluSerLysLeu...AsnTyrLysTrpSerProArgCysSerProArgLeuArgSer
 ThrAlaAspProTrpThrGlyLeuLeuAlaHisAspLeuMETLeuMETThrSerLysAlaProLeuLeuArgLys
 SerGlnLeuHisAsnLeuTyrTyrAlaProIleGlnGlnGluAlaValArgAlaValValGlyGlnProProGln
 GlnHisLeuGlyPheProValGluMETGly

(2) INFORMATIONS POUR LA SEQ ID NO: 29: protéine env (cadre de lecture 3)

(i) CARACTERISTIQUES DE LA SEQUENCE:

(B) TYPE: acide aminé,

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéine

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 29:

AlaProSerGlyGluGlnArgThrGlyHisTyrProGlyGluAspTrpGlnLeuIleLeuProThrSerProAsn
 LeuArgAspPheSerIleTyr...SerGly...IleLeuSerArgValGlyGlnArgProSerProValGlyGln
 LysArgProLysArg.....ArgHis...PheMETLys...PheProAspSerAspPheProGluAlaTyrArg
 ValThrIleAlaLeuLeuSerArgProGln...ProArgGluTyrProArgArg...ValTyrAspIleThrTyr
 ThrAlaProGluGlyHisSerProGlnGlyArgSerArgLys...METLysHisSerLysAspIle...LysSer
 LysProArgLysProThrSerHisGlyLeuLeuCysCysLeu...Pro...LysGluSerAlaThrPheProLys
 LysGlnAspLeuAlaHisThrLysCysCysMETGluGlyProSer...ProMETThrLeuCysLeuThrGlnAsp
 SerGlnLeuSerCysArgHisHisLeuLeuSerGlnIleSerThrSerSer...AsnIleThrArgAsnLeuSer
 LeuArgArgGlyLysArgThrIleProProLeu...HisGlyIleSerGlnValProSerLeu...PheProIle

ProArgTyrIleLeuGlyArgThrLeuProSerHisPheIleTyrProAsnCysGly...SerGlyTrpSerGly
 ValLeuAspThrSerHisLeuSerGlnIleLeuAspThrAlaLysGlyThr...LysSerArgArgGlnArg...
 LeuPheLeu...ThrSerArgGlyPheAlaProAlaLeuGlnThrThrThrArgArgLysValThrLysIleIle
 AsnProHisGlyProProLeuSerTyrPheSerLeuTyrCysSerPheThrLeuPheHisSerHisCysThrPro
 SerMETProLeuTyrAspGln...LeuProLeuProArgValSerMETGluAsnAlaAlaSerArgLysTyr...
 CysProIleVal...GluSerPhe...GlyAsnProHisLeuHisCysProHisProTyrAlaProGlnLeuLeu
 SerLeuCysHisSerLeuHisAlaCysLysTyrSerLeuLeuAspArgLysAsnAsp...Ser...LeuSerTrp
 ArgThrTrpSerHisCysLeuLeuAspLeuLeuHisProAsnTrpTyrVal...TrpGlyTrpSerSerArgSer
 GlyLysArgLysThrCysLysArgSerAsnLeuProThrHisProGlyThrTrpHisLeu...ProLeuGlnArg
 ThrArgSerLeuLysThrThr...AsnProProTyrProTyrSerProGlyLysProIle...TyrHisProHis
 TrpAlaPro...GlyLeuGlyProLysProTyr...LeuLeuAspMETProProProGluLeuGlnAlaIleCys
 PheAsnProCysThr...ThrMETGluGlnLeuGlnHisArgAsnLysHisHisPheArgPheSerArgThrSer
 CysPheGlnSerGlyAsnAsnProTyrLeuLysProHisLeuCysLysIle...GlnTyrTyrIleHisAsnGln
 LeuProMETHisGlnValGlyAsnSerSerHisThrAsnSerLeuProThrLeuArgAsnIlePheCysLeuTrp
 TyrLeuSerLeuSerLeuPheGluTrpLeuPheArgIleTyrValLeuProLeuIleLeuSerAlaProTyrAsp
 HisLeuHis...ThrArgPheIleGlnLeuCysHisIle...AlaProGlnGlnLysSerThrHisSerSerPhe
 CysTyrArgSerArgSerAlaArgCysThrArgTyrTrpHisTrpArgTyrHisAsnLeuTyrSerValLeuLeu
 GlnThrIleSerArgThrLysTrpGlyHisGlyThrGlyArgArgLeuProGlyHisLeuAlaArgSerThr...
 LeuProSerSerSerSerProSerLysSerLysSerPheArgLeuAlaAsnArg...LysArgGlyAsnLeuPhe
 IlePheArgGlyArgMETLeuLeuLeuCys...SerIleArgAsnArgHis...GluSer...ArgAsnSerArg
 SerAsnThrThr...SerArgGlyAlaSerLysHisTrpThrLeuGlyProProGlnProMETAspAlaLeuAsp
 SerProLeuLeuArgThrSerSerSerTyrAsnIleAlaThrProLeuTrpThrLeuTyrLeu...ProProCys
 ...LeuCysLeuPheGlnAsnArgSerCysLysThrThrAsnGlyAlaGlnAspAlaValGlnAsp...AspLeu
 ProGlnThrProGlyProAlaCys...ProThrIle...Cys.....HisGlnArgHisProSer...GlyAsn
 LeuSerCysThrThrSerThrThrProGlnPheSerArgLysGlnLeuGluArgSerSerAlaAsnLeuProAsn

SerThr...ValPheLeuLeuArgTrpGly

(2) INFORMATIONS POUR LA SEQ ID NO:30 : G1F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 26 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:30 :

GGACCATAGAGGACACTCCAGGACTA

(2) INFORMATIONS POUR LA SEQ ID NO:31 : G1R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 25 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:31 :

CCTCAGTCCTGCTGCTGGATCATCT

(2) INFORMATIONS POUR LA SEQ ID NO:32 : G2F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:32 :

CCTCCAAGCAGTGGGAGGAAGAGAATT

(2) INFORMATIONS POUR LA SEQ ID NO:33 : G2R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 28 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:33 :

CCTTCCCTGTGTTATTGTGGACATCATT

(2) INFORMATIONS POUR LA SEQ ID NO:34 : G4F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 30 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:34 :

GGAAGAAGTCTATGAATTATTCAATGATGT

(2) INFORMATIONS POUR LA SEQ ID NO:35 : G3F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:35 :

GGGACACAGAATCAGAACATGGAGATT

(2) INFORMATIONS POUR LA SEQ ID NO:36 : G4R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:36 :

GCCTTCAGAAGAGTCAGGTGACAGAGA

(2) INFORMATIONS POUR LA SEQ ID NO:37 : G5R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 25 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:37 :

GAGCCTCCAAAGTCCACTTGCCTGA

(2) INFORMATIONS POUR LA SEQ ID NO:38 : E1F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 29 paires de bases
 - (B) TYPE: nucléotide

- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéair

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:38 :

GATTTCA GTATCTACTAGTCTGGGTAGAT

(2) INFORMATIONS POUR LA SEQ ID NO:39 : E1R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:39 :

CTAGGAAATCCAGCTAGTCCTGTCTCA

(2) INFORMATIONS POUR LA SEQ ID NO:40 : E2F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 28 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:40 :

CCAAGACAGCCAACTTAGTTGCAGACAT

(2) INFORMATIONS POUR LA SEQ ID NO:41 : E2R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 28 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:41 :

GGACGCTGCATTCTCCATAGAAACTCTT

(2) INFORMATIONS POUR LA SEQ ID NO:42 : E3F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 29 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:42 :

GCAATACTACATACACAACCAACTCCCAA

(2) INFORMATIONS POUR LA SEQ ID NO:43 : E3R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 26 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:43 :

GGGGGAGGCATATCCAACAGTTAGTA

(2) INFORMATIONS POUR LA SEQ ID NO:44 : E4F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 30 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:44 :

CCATCTACTGAACAAGATTTATACACTT

(2) INFORMATIONS POUR LA SEQ ID NO:45 : E4R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 28 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:45 :

AATGCCAGTACCTAGTGCACCTAGCACT

(2) INFORMATIONS POUR LA SEQ ID NO:46 : E5F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 31 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:46 :

CGAATACAACGTAGAGCAGAGGAGCTTCGAA

(2) INFORMATIONS POUR LA SEQ ID NO:47 : E6F

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 28 pairs de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADnc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:47 :

AGCCCAAGATGCAGTCCAAGACTAAGAT

(2) INFORMATIONS POUR LA SEQ ID NO:48 : E5R

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADnc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:48 :

GCGTAGTAGAGGTTGTGCAGCTGAGAT

(2) INFORMATIONS POUR LA SEQ ID NO:49 : ExF

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADnc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:49 :

CCCTTACCAAGAGTTTCTATGGAGAAT

(2) INFORMATIONS POUR LA SEQ ID NO:50 : ExR

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 27 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRINS: simple
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADnc (amorce)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO:50 :

ACCGCTCTAACTGCTTCCTGCTGAATT

(2) INFORMATIONS POUR LA SEQ ID NO: 51: protéine gag

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (B) TYPE: acide aminé,
 - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: protéin
(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 51:

TSFVEKANGVKCHKYKLSFHKETTHNYVKSIVIYALQEAFRVYLPILPASPTPSPTNKDPPSTQMVQKEIDKRVN
SEPKSANIPQLXPLQAVGGREFGPARVHVFPFSLPDLKQIKTDLGKFSDNPDGYIDVLQGLGQFFDLTWRDIMSL
LNQTLTPNERSATITAAXEFGDLWYLSQVNDRMTTEEREXFPTGQQAVPSLDPHWDTESEHGDWCCRHLITCVL
EGLRKTRKKSMNYSMSTITQGREENPTAFLERLREALRKASLSPDSSEGQLILKRKFITQSAADIRKKLQKS
AVGPEQNLETLLNLATSVFYNRDQEEQAEQDKRDXXXGHRFSDHPQASGLWRLWKREKLKLNAXXGLLPVRST
RTLXKRLSKXXAAPSMSPLISRESLEGPLPQGTKVLXVRSHXPD/SSRT

REVENDEICATIONS

1°) Fragment d'acide nucléique purifié, caractérisé en ce qu'il comprend tout ou partie d'une séquence codant pour une séquence rétrovirale endogène humaine, qui présente au moins des motifs rétroviraux de type *env*, répondant à la séquence SEQ ID NO:1 ou à une séquence présentant un niveau d'homologie avec ladite séquence SEQ ID NO:1 supérieur ou égal à 80% sur plus de 190 nucléotides ou supérieur ou égal à 70% sur plus de 600 nucléotides pour les domaines de type *env*.

2°) Fragment d'acide nucléique selon la revendication 1, caractérisé en ce qu'il présente à la fois de motifs rétroviraux correspondant à un domaine *env* et répondant à la séquence SEQ ID NO:1 et des motifs rétroviraux correspondant à un domaine *gag* et répondant à la séquence SEQ ID NO:2 ou à une séquence présentant un niveau d'homologie supérieur ou égal à 80% sur plus de 190 nucléotides ou supérieur ou égal à 70% sur plus de 600 nucléotides pour les domaines de type *env* et un niveau d'homologie supérieur ou égal à 90% sur plus de 700 nucléotides ou supérieur ou égal à 70% sur plus de 1200 nucléotides pour les domaines de type *gag*, lesquels motifs ne présentent aucune insertion ou délétion supérieure à 200 nucléotides.

3°) Fragment d'acide nucléique, caractérisé en ce qu'il comprend un segment d'une séquence selon la revendication 1 ou la revendication 2 et notamment les séquence SEQ ID NO:3-24, les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes ainsi que les fragments issus des régions codantes des séquences précédentes correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires.

4°) Transcrits, caractérisé en ce qu'ils sont générés à partir des séquences selon l'une quelconque des revendications 1 à 3.

5°) Réactif de diagnostic pour la détection différentielle de séquences nucléiques endogènes humaines complètes ou partielles, présentant des motifs rétroviraux, sélectionnés parmi les séquences SEQ ID NO:1 et/ou SEQ ID NO:2, caractérisé en ce qu'il est sélectionné dans le groupe constitué par les séquences SEQ ID NO:1-50, les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes, par les fragments nucléotidiques capables de définir ou d'identifier les séquences SEQ ID NO:1 et/ou SEQ ID NO:2 et

toute séquence flanquante ou les chevauchants ainsi que par les fragments issus des régions codantes des séquences SEQ ID NO:1-24, correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires, éventuellement marquées avec un marqueur approprié.

5 6°) Réactif selon la revendication 5, caractérisé en ce qu'il est choisi dans les régions situées entre les nucléotides 3065 et 4390 et les nucléotides 6965 et 9550 de la SEQ ID NO:3.

 7°) Réactif selon la revendication 5, caractérisé en ce qu'il est sélectionné parmi les séquences SEQ ID NO:30-50 et en ce qu'il est apte à être utilisé
10 comme amorce.

 8°) Réactif selon la revendication 5, caractérisé en ce qu'il est sélectionné parmi les séquences suivantes :

 - un fragment de 1505 nt amplifié par le couple d'amorces SEQ ID NO:30 et SEQ ID NO:31 (amorces G1F et G1R),
15 - un fragment de 2529 nt amplifié par le couple d'amorces SEQ ID NO:38 et SEQ ID NO:39 (amorces E1F et E1R) et en ce qu'il est apte à être utilisé comme sonde.

 9°) Procédé de détection rapide et différentiel des séquences nucléiques rétrovirales endogènes de type *env* ou *env* et *gag*, de leurs variants
20 normaux ou pathologiques, par hybridation et/ou amplification génique, réalisé à partir d'un échantillon biologique, lequel procédé est caractérisé en ce qu'il comprend :

 (a) une étape dans laquelle l'on met en contact un échantillon biologique à analyser avec au moins une sonde selon la revendication 5, la revendication 6 ou la revendication 8 et

25 (b) une étape dans laquelle on détecte par tout moyen approprié le ou les produits résultants de l'interaction séquence nucléotidique-sonde.

 10°) Procédé de détection selon la revendication 9, caractérisé en ce qu'il comprend :

 * préalablement à l'étape (a) :
30 . une étape de préparation du tissu ou du liquide biologique concerné,

. une étape d'extraction de l'acide nucléique à détecter, et
 . au moins un cycle d'amplification génique mis en œuvre à l'aide
 d'au moins un réactif selon l'une quelconque des revendications 5 à 7 et

* postérieurement à l'étape (b) :

5 . une étape de comparaison des séquences nucléiques obtenues dans
 ledit échantillon biologique avec les séquences rétroviraux endogènes humaines selon
 l'une quelconque des revendications 1 à 3, par tout moyen approprié et notamment par
 séquençage, Southern-blot, coupure de restriction, SSCP ou toute autre méthode
 permettant d'identifier une insertion ou une délétion ou encore une simple mutation
 10 entre les différentes séquences comparées.

11°) Procédé de détection des transcrits selon la revendication 4,
 caractérisé en ce qu'il comprend :

- le prélèvement des ARN messagers provenant de tissus témoins et
 de tissus prélevé chez des patients et

15 - l'analyse qualitative et/ou quantitative desdits ARNm, par hybri-
 dation *in situ*, par dot-blot, Northern-blot, RNase mapping ou RT-PCR, à l'aide d'un
 réactif de diagnostic selon l'une quelconque des revendications 5 à 8.

12°) Produits de traduction, caractérisés en ce qu'ils sont codés par
 une séquence nucléotidique selon l'une quelconque des revendications 1 à 3.

20 13°) Peptide, caractérisé en ce qu'il est susceptible d'être exprimé à
 l'aide d'une séquence nucléotidique sélectionnée dans le groupe constitué par les
 séquences SEQ ID NO:1-24 selon l'une quelconque des revendications 1 à 3, selon les
 combinaisons offertes par l'usage des différents cadres de lecture possibles.

25 14°) Peptide selon la revendication 13, caractérisé en ce qu'il
 englobe les peptides dérivés comprenant entre 5 et 540 aminoacides.

15°) Peptide selon la revendication 13 ou la revendication 14,
 caractérisé en ce qu'il est sélectionné parmi les séquences SEQ ID NO :25-29 et la
 séquence SEQ ID NO :51.

30 16°) Peptide selon l'une quelconque des revendications 13 à 15,
 caractérisé en ce qu'il est obtenu à partir des séquences nucléiques selon l'une
 quelconque des revendications 1 à 3, dans lesquelles au moins un codon non-sens peut

être remplacé par un codon codant pour l'un des aminoacides suivants : Phe (F), Leu (L), Ser (S), Tyr (Y), Cys (C), Trp (W), Gln (Q), Arg (R), Lys (K), Glu (E) ou Gly (G).

17°) Anticorps, caractérisé en ce qu'il est dirigé contre l'un ou
5 plusieurs des peptides selon l'une quelconque des revendications 13 à 16.

18°) Procédé de dépistage immunologique différentiel de séquences
rétrovirales endogènes humaines de la famille HERV-7q normales ou pathologiques,
caractérisé en ce qu'il comprend la mise en contact d'un échantillon biologique avec
un anticorps selon la revendication 17, la lecture du résultat étant révélée par un
10 moyen approprié, notamment EIA, ELISA, RIA, fluorescence.

19°) Procédé d'identification et de détection de motifs rétroviraux
endogènes, anormalement exprimés dans le cadre de pathologies associées au cancer,
ou de neuropathologies en particulier autoimmunes, au premier rang desquelles la
sclérose en plaques, caractérisé en ce qu'il comprend l'analyse comparée des séquences
15 extraites d'un échantillon biologique avec les séquences selon l'une quelconque des
revendications 12 à 16.

18°) Application des séquences selon l'une quelconque des revendi-
cations 1 à 6 ou 12 à 16 au diagnostic, au pronostic, à l'évaluation de la susceptibilité
génétique, à toutes maladies humaines induites, innées ou acquises en particulier
20 celles à composantes cancéreuses, autoimmunes et/ou à incidence neurologique,
comme la sclérose en plaques, les syndromes associés et les maladies neurodégénéra-
tives où intervient tout ou partie des séquences selon l'une quelconque des revendica-
tions 1 à 5 et des formes endogènes ou exogènes apparentées.

19°) Séquences nucléiques hybrides, caractérisées en ce qu'elles
25 comprennent des séquences ou motifs selon l'une quelconque des revendications 1 à
5, combinés avec des séquences ou motifs d'origine endogène ou d'origine ou induits
de manière exogène.

20°) Vecteur recombinant de clonage ou d'expression, caractérisé en
ce qu'il comprend une séquence nucléique selon l'une quelconque des revendications
30 1 à 4.

de la RNase H,

- les régions *gag* et *pol* pourraient être considérées comme jointives avec un passage de la région *gag* à la région *pol* par un décalage du cadre de lecture.

La présente invention englobe les séquences appartenant à la famille

- 5 HERV-7q telle que définie ci-dessus (présence de la séquence SEQ ID NO:1 ou d'une séquence homologue ou présence à la fois des séquences SEQ ID NO:1 et SEQ ID NO:2) et notamment les séquences SEQ ID NO:3-24 ; elle englobe également les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes ainsi que les fragments issus des régions codantes des
- 10 séquences précédentes correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires. (SEQ ID NO :30-50)

Ces différents fragments peuvent avantageusement être utilisés comme amorces ou comme sondes ; ils s'hybrident spécifiquement à une séquence de la famille HERV-7q.

- 15 Parmi ces fragments, on peut citer, de préférence les fragments suivants:

- un fragment de 182 nucléotides répété deux fois, situé en amont du domaine *gag* aux positions 2502-2611/2613-2865 de la SEQ ID NO:3 ;

Amorces et sondes spécifiques de la région *gag*

- 20 - une amorce G1F, sens, localisée dans la région amont du domaine *gag* de HERV-7q : 5' GGACCATAGAGGACACTCCAGGACTA 3' (SEQ ID NO:30);

- une amorce G1R, anti-sens, localisée dans la région 3' terminale du domaine *gag* : 5' CCTCAGTCCTGCTGCTGGATCATCT 3' (SEQ ID NO :31)

- 25 - le fragment de 1505 nt amplifié par le couple G1F-G1R est utilisé afin de générer les sondes aptes à hybrider les différents produits d'amplification des PCR ;

- une amorce G2F, sens nichée : (SEQ ID NO :32)

5' CCTCCAAGCAGTGGGAGGAAGAGAATT 3'

REVENDEICATIONS

1°) Fragment d'acide nucléique purifié, caractérisé en ce qu'il comprend tout ou partie d'une séquence codant pour une séquence rétrovirale endogène humaine, qui présente au moins des motifs rétroviraux de type *env*, répondant à la séquence SEQ ID NO:1 ou à une séquence présentant un niveau d'homologie avec ladite séquence SEQ ID NO:1 supérieur ou égal à 80% sur plus de 190 nucléotides ou supérieur ou égal à 70% sur plus de 600 nucléotides pour les domaines de type *env*.

2°) Fragment d'acide nucléique selon la revendication 1, caractérisé en ce qu'il présente à la fois de motifs rétroviraux correspondant à un domaine *env* et répondant à la séquence SEQ ID NO:1 et des motifs rétroviraux correspondant à un domaine *gag* et répondant à la séquence SEQ ID NO:2 ou à une séquence présentant un niveau d'homologie supérieur ou égal à 80% sur plus de 190 nucléotides ou supérieur ou égal à 70% sur plus de 600 nucléotides pour les domaines de type *env* et un niveau d'homologie supérieur ou égal à 90% sur plus de 700 nucléotides ou supérieur ou égal à 70% sur plus de 1200 nucléotides pour les domaines de type *gag*, lesquels motifs ne présentent aucune insertion ou délétion supérieure à 200 nucléotides.

3°) Fragment d'acide nucléique, caractérisé en ce qu'il comprend un segment d'une séquence selon la revendication 1 ou la revendication 2 et notamment les séquences SEQ ID NO:3-24, les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes ainsi que les fragments issus des régions codantes des séquences précédentes correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires.

4°) Transcrits, caractérisé en ce qu'ils sont générés à partir des séquences selon l'une quelconque des revendications 1 à 3.

5°) Réactif de diagnostic pour la détection différentielle de séquences nucléiques endogènes humaines complètes ou partielles, présentant des motifs rétroviraux, sélectionnés parmi les séquences SEQ ID NO:1 et/ou SEQ ID NO:2, caractérisé en ce qu'il est sélectionné dans le groupe constitué par les séquences SEQ ID NO:1-50, les séquences nucléiques complémentaires et les séquences inverses complémentaires des séquences précédentes, par les fragments nucléotidiques capables de définir ou d'identifier les séquences SEQ ID NO:1 et/ou SEQ ID NO:2 et

toute séquence flanquante ou les chevauchants ainsi que par les fragments issus des régions codantes des séquences SEQ ID NO:1-24, correspondant à un cadre glissant supérieur ou égal à 14 nucléotides ou leurs séquences complémentaires, éventuellement marquées avec un marqueur approprié.

5 6°) Réactif selon la revendication 5, caractérisé en ce qu'il est choisi dans les régions situées entre les nucléotides 3065 et 4390 et les nucléotides 6965 et 9550 de la SEQ ID NO:3.

 7°) Réactif selon la revendication 5, caractérisé en ce qu'il est sélectionné parmi les séquences SEQ ID NO:30-50 et en ce qu'il est apte à être utilisé
10 comme amorce.

 8°) Réactif selon la revendication 5, caractérisé en ce qu'il est sélectionné parmi les séquences suivantes :

 - un fragment de 1505 nt amplifié par le couple d'amorces SEQ ID NO:30 et SEQ ID NO:31 (amorces G1F et G1R),
15 - un fragment de 2529 nt amplifié par le couple d'amorces SEQ ID NO:38 et SEQ ID NO:39 (amorces E1F et E1R) et en ce qu'il est apte à être utilisé comme sonde.

 9°) Procédé de détection rapide et différentiel des séquences nucléiques rétrovirales endogènes de type *env* ou *env* et *gag*, de leurs variants
20 normaux ou pathologiques, par hybridation et/ou amplification génique, réalisé à partir d'un échantillon biologique, lequel procédé est caractérisé en ce qu'il comprend :

 (a) une étape dans laquelle l'on met en contact un échantillon biologique à analyser avec au moins une sonde selon la revendication 5, la revendication 6 ou la revendication 8 et

25 (b) une étape dans laquelle on détecte par tout moyen approprié le ou les produits résultants de l'interaction séquence nucléotidique-sonde.

 10°) Procédé de détection selon la revendication 9, caractérisé en ce qu'il comprend :

 préalablement à l'étape (a) :

30 . une étape de préparation du tissu ou du liquide biologique concerné,

. une étape d'extraction de l'acide nucléique à détecter, et
. au moins un cycle d'amplification génique mis en œuvre à l'aide
d'au moins un réactif selon l'une quelconque des revendications 5 à 7 et

* postérieurement à l'étape (b) :

5 . une étape de comparaison des séquences nucléiques obtenues dans
ledit échantillon biologique avec les séquences rétrovirales endogènes humaines selon
l'une quelconque des revendications 1 à 3, par tout moyen approprié et notamment par
séquençage, Southern-blot, coupure de restriction, SSCP ou toute autre méthode
permettant d'identifier une insertion ou une délétion ou encore une simple mutation
10 entre les différentes séquences comparées.

11°) Procédé de détection des transcrits selon la revendication 4,
caractérisé en ce qu'il comprend :

- le prélèvement des ARN messagers provenant de tissus témoins et
de tissus prélevé chez des patients et

15 - l'analyse qualitative et/ou quantitative desdits ARNm, par hybri-
dation *in situ*, par dot-blot, Northern-blot, RNase mapping ou RT-PCR, à l'aide d'un
réactif de diagnostic selon l'une quelconque des revendications 5 à 8.

12°) Produits de traduction, caractérisés en ce qu'ils sont codés par
une séquence nucléotidique selon l'une quelconque des revendications 1 à 3.

20 13°) Peptide, caractérisé en ce qu'il est susceptible d'être exprimé à
l'aide d'une séquence nucléotidique sélectionnée dans le groupe constitué par les
séquences SEQ ID NO:1-24 selon l'une quelconque des revendications 1 à 3, selon les
combinaisons offertes par l'usage des différents cadres de lecture possibles.

25 14°) Peptide selon la revendication 13, caractérisé en ce qu'il
englobe les peptides dérivés comprenant entre 5 et 540 aminoacides.

15°) Peptide selon la revendication 13 ou la revendication 14,
caractérisé en ce qu'il est sélectionné parmi les séquences SEQ ID NO :25-29 et la
séquence SEQ ID NO :51.

30 16°) Peptide selon l'une quelconque des revendications 13 à 15,
caractérisé en ce qu'il est obtenu à partir des séquences nucléiques selon l'une
quelconque des revendications 1 à 3, dans lesquelles au moins un codon non-sens peut

être remplacé par un codon codant pour l'un des aminoacides suivants : Phe (F), Leu (L), Ser (S), Tyr (Y), Cys (C), Trp (W), Gln (Q), Arg (R), Lys (K), Glu (E) ou Gly (G).

17°) Anticorps, caractérisé en ce qu'il est dirigé contre l'un ou
5 plusieurs des peptides selon l'une quelconque des revendications 13 à 16.

18°) Procédé de dépistage immunologique différentiel de séquences
rétrovirales endogènes humaines de la famille HERV-7q normales ou pathologiques,
caractérisé en ce qu'il comprend la mise en contact d'un échantillon biologique avec
un anticorps selon la revendication 17, la lecture du résultat étant révélée par un
10 moyen approprié, notamment EIA, ELISA, RIA, fluorescence.

19°) Procédé d'identification et de détection de motifs rétroviraux
endogènes, anormalement exprimés dans le cadre de pathologies associées au cancer,
ou de neuropathologies en particulier autoimmunes, au premier rang desquelles la
sclérose en plaques, caractérisé en ce qu'il comprend l'analyse comparée des séquences
15 extraites d'un échantillon biologique avec les séquences selon l'une quelconque des
revendications 12 à 16.

20°) Application des séquences selon l'une quelconque des revendi-
cations 1 à 6 ou 12 à 16 au diagnostic, au pronostic, à l'évaluation de la susceptibilité
génétique, à toutes maladies humaines induites, innées ou acquises en particulier
20 celles à composantes cancéreuses, autoimmunes et/ou à incidence neurologique,
comme la sclérose en plaques, les syndromes associés et les maladies neurodégénéra-
tives où intervient tout ou partie des séquences selon l'une quelconque des revendi-
cations 1 à 5 et des formes endogènes ou exogènes apparentées.

21°) Séquences nucléiques hybrides, caractérisées en ce qu'elles
25 comprennent des séquences ou motifs selon l'une quelconque des revendications 1 à
5, combinés avec des séquences ou motifs d'origine endogène ou d'origine ou induits
de manière exogène.

22°) Vecteur recombinant de clonage ou d'expression, caractérisé en
ce qu'il comprend une séquence nucléique selon l'une quelconque des revendications
30 1 à 4.

CCCTGGGGCGGGCTTCTCTTCTGGGATGAGGGCAAACGCCCTGGAGATACAGCAATTATCTTGCAACTGAG	71	
AGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACCACGTCAC	143	
CTTTAAACACGGGGCTTGCAACTTAGGCTACACCTGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCA	215	
AAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGGAGGAGGACAACAAATCGGGATA	287	
TAAACCCAGGCATTGAGCTGGCAACAGCAGCCCCCTTTGGGTCCCTTCCCTTTGTATGGGAGCTGTTTC	359	région
ATGCTATTTCACTCTATTAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGGCTCGAGCTGAGCT	431	répétée
TTTGCTCACCGTCCACCACTGCTGTTTGCCACCACCGCAGACCTGCCGCTGACTCCCATCCCTCTGGATCCT	503	R1
GCAGGGTGTCCGCTGTGCTCTGATCCAGCGAGGCGCCCATTTGCCGCTCCCAATTGGGCTAAAGGCTTGCCA	575	
TTGTTCTGCACGGCTAAGTGCTGGGTTTGTCTAATTGAGCTGAACACTAGTCACTGGGTTCCATGGTTC	647	
TCCTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACATGGCCCAAGATTCCATTCCTTGAAT	719	
CCGTGAGGCCAAGAATCCAGGTCAGAGAATACGAGGCTTGCCACCATCTTGAAGCGGCTGTACCATCT	791	
TGGAAGTGGTTCACCACTCTTGGGAGCTCTGTGAGCAAGGACCCCCGGTAACATTTTGGCAACCACGAA	863	
CGGACATCCAAAGTGGTGAGTAATATTGGACCACTTTCACTTGTCTATTCTGCTCTATCCTTCTTAGAATTG	935	
GAGGAAAAATACCGGCACTTGTGCGCCAGTTAAAAACGATTAGTGTGGCCACCGGACTTAAGACTCAGGTGT	1007	
GAGGCTATCTGGGGAAGGGCTTCTAACAACCCCCAACCTTCTGGGTTGGGGACTTGGTTTGCTCAAGCC	1079	
AGCTTCCACTTTCAGTTTCTTGGGGAAGCCGAGGGCCGACTAGAGGCAGAAAGCTGTCGCTCTGAACCTCCC	1151	
GGCAGTAGCCGGTTGAGATCATGGTGTAGCCAGAAGTCTCAACAGTCCGCTATGCATGCACCCCTATCTTC	1223	
CTTCTGACCCATACCTCCTGGGTCGCCAACAACCTTTCTTCAAGTGTAGCCCCAAAATTTCTCTTACCTC	1295	
TGAATATACTTCTCTGATCCCTGCCTCCTAGGTACTATTGGTTGAGACTTCCATTCTCTAGCAAGTTGT	1367	
ATCTCCAAAGGGATCTAAGGAAGCTCTGCGCTGCGTCTTAGGCACCTAGGCTATAACCCAGGGAGTCTTAT	1439	
CCCTGGTGTCCCTCCCAATTTAGGCATACAGCTCTTGACATGGGCAGTTATGTAGGACCCACTCCCCACCAC	1511	
CCTTGCCAGGGCCCCAAGTTTGTAAATGGCTGAGGGAAGAGAGACAGAGGAGAGAGAGAGAAATGGAGGA	1583	
GAAAGAGAGAGAGACAGAGAGAGAGAGAGACAGTGAAGAGACAGAAGAGAGAGAGACAAAGAGGAGAG	1655	
AGAGAGAGTCAAGAGAGAAAGAGAAAGAGAAATAGTAAAAACAGTGTGCCCTATTCTTTTAAAGGCCA	1727	
GGGTAAATTTAAACCTGTACTTGATAATTGAAGGTCTTCTGTGACCCATAGCACTCCAATCCACTTTG	1799	
TGGTCAGTGTAAATAGAGCATAGGCCGAAAGCACTGAGGCCATTGCAACCCGCTAGCTTCCCTATCAAAA	1871	
TCCTTAACCCAGTAACCCGAGATGGACCAATGCATTGCTCGGTAGCGCAACTGCTTTGTCTAAAGTAGA	1943	
AAAGTAACCTTTAGAGGAAACCTCATTGTGAGCACACCTCACCTGTCAGAAATTATCTAATAAAAAAGCA	2015	
AAAAGGTATCTTACTAATCAAAAATCTTAAAGTATGGGGCTATTCTGTTAGAAAAAGGTAAATGTAACCTCA	2087	
ACCACTGATAATTCCTTAAACCAGCAGAGTTCTAACCGGATTTAAATCTTAATTACCATACAAAGTCCG	2159	
ACCAGACCTAGGCGGAACCTCCCTCAGGACAGGACGATAGATGGTTCCTCCAGGTGATTGAGGAAAAAAAC	2231	
CACAATGGGTATTGCTAATTGATACGGGGACTCTTGTGGAAGCAGAGTTAGAAAAATTGCCTAATAACTGG	2303	
TCCTCTCAACCGTGTGAGCTGTTGCACTCAGCCAAGCTTAAAGTACTTACAGAATCAAAAGACTATCTCA	2375	
ATCCTGATTCAAAAGGTTAGCTACACCCCTCTCTGTAATGCATTTGCATAAGAACTTGTATGGGAATGCAT	2447	
CTTGATGGGCGAGCTGGGTTGTTATAAAATAGGAACCCAGCCAGCTCTAGGACTCACCCCTGAGCGCAAG	2519	
GCAATGTTGGGCATGTGGTAAAGGACCACTAGAATCCAGCGCCAGCCACTTCTTTGTGGTCAAGAAA	2591	régions
GGCGGGAAGAGGGGTGCAGGACTGCTACATCGGTAAGCATAACTATCCGATAAAGAGAGGTCCATGGGTGG	2663	répétées
TTACGCACTCTGGAAGGAACCTACCCCTGAGCACAAAGGCAATGTTGGGCACGCTGGTAAGGACCACTAG	2735	en tandem
AAATCCAGCAGCTGGACCCCTTCTTTGTGGTCAAGAGAGGAGGAAACAGGTGCAAGCATCAG	2807	R2
TGAGCATACTAATTTCGATAAGCAGAGGTCCATGGGTGGTGTATGCACCTGGAAGAAATAAGCATTAGGACC	2879	
ATAGAGGACACTCCAGGACTAAGGCTCATCGGAAAATGACTAGGGTGTGCTGGCATCCCTATGTTCTTTTTC	2951	
AGATGGGAAACGTTCCCGCAAGACAAAACGCCCTAAGACGTTATTCTGGAGAATTGGGACCAATTGACC	3023	
CTCAGACACTAAGAAAGAAACGACTTATATTCTTCTGCACTGCCGCTGGCACTCTGAGGGAAGTATAAAT	3095	
TATAACACCATCTTACAGCTAGACCTCTTTTGTAGAAAAGGCAAAATGGAGTGAAGTGCCATAAGTACAACT	3167	
TTCTTTTCATTAAGAGACAACCTCACAATTATGTAAAAAGTGTGATTATGCCCCTACAGGAAGCCTTCAGAGT	3239	
CTACCTCCCTATCCCAGCATCCCCGACTCCTTCCCCAATAAAGGACCCCTTCAACCCAAATGGTCCA	3311	
AAAGGAGATAGACAAAAGGGTAAACAGTGAACCAAGAGTGCCAAATATTCCCCAATTATGACCCCTCCAAGC	3383	
AGTGGGAGGAAGAGAAATTCGGCCAGCCAGAGTGCATGTGCTTTTCTCTCCAGACTTAAAGCAAAATAA	3455	
AACAGACTTAGGTAATTTCTCAGATAACCTGTATGGCTATATTGATGTTTTACAAGGTTAGGACAATTCTT	3527	
TGATCTGACATGGAGAGATATAATGTCACTGCTAAATCAGACACTAACCCCAATAGAGAAAGTGCCACCAT	3599	
AACTGCAGCCTGAGAGTTTGGCGATCTCTGGTATCTCAGTCAGGTCAATGATAGGATGACAACAGAGGAAAG	3671	domaine
AGAATGATTCCCCACAGGCCAGCAGGCACTTCCAGTCTAGACCCCTCATTTGGGACACAGAATCAGAACATGG	3743	gaq
AGATTGGTGTGTCAGACATTGCTAATCTTGTGTGTGCTAGAAAGGACTAAGGAAACTAGGAAGAAGTCTATGAA	3815	
TTACTCAATGATGTCCACCATAACACAGGGAAGGGAAGAAATCTACTGCCTTTCTGGAGAGACTAAGGGA	3887	
GGCATTGAGGAAGCGTGCTCTGTGACCTGACTCTTCTGAAGGCCAACTAATCTTAAAGCGTAAGTTTAT	3959	
CACCTCAGTCAGCTGCAGACATTAGAAAAAACTTCAAAAGTCTGCCGTAGGCCCGGAGCAAAACTTAGAAAC	4031	
CCTATTGAACCTTGGCAACCTCGGTTTTTATAATAGAGATCAGGAGGAGCAGGCGGAACAGGACAAACGGGA	4103	
TTAAAAAAAGGCCACCGCTTTAGTCATGACCCCTCAGGCAAGTGGACTTTGGAGGCTCTGGAAGGGGAAAA	4175	
GCTGGGCAAAATGAATGCCTAATAGGGCTTGCTTCCAGTGGGTCTACAAGGACACTTTAAAAAAGATTGTC	4247	
CAAGTAGAAGTAAAGCCGCCCTCGTCCATGCCCTTATTTCAGGGAATCACTGGAAGGCCCACTGCCCA	4319	
GGGGACAAAGGTCTCTGAGTCAGAAGCCACTAACAGATGATCCAGCAGCAGGACTGAGGGTGCCTGGGGC	4391	
AAGCGCCATCCCATGCCATCACCTCAGAGCCCTGGGTATGCTTGACCATTGAGGGCCAGGAGGTTGTCT	4463	
CCTGGACACTGGTGCGGTCTTCTAGTCTTACTCTTCTGTCCCGGACAACTGCTCCAGATCTGTCACTAT	4535	
CTGAGGGGCTCCTAAGACGGGCAGTCACTAGATACTTCTCCAGCCACTAAGTTATGACTGGGGAGCTTTAT	4607	
TCTTTTACATGCTTTTCTAATTATGCTTGAAGCCCCACTACCTTGTAGGGAGAGACATTCTAGCAAAAG	4679	
CAGGGGCCATTATACAGTGAACAGAGGACAATATGGACGAGCAAGAAATGCCGCTCCTGTTCAAGTTAACTAA	4751	
ATCCTGAAGTCTGGGCAACAGAGGACAATATGGACGAGCAAGAAATGCCGCTCCTGTTCAAGTTAACTAA	4823	
AGGATTCACCTCCTTTCCCTACCAAGGCAGTACCCCTCAGACCCAAGGCCCAACAGGACTCCAAAGA	4895	
TTGTTAAGGACCTAAGAGCCCAAGGCTAGTAAACCATGCAGTAACCCCTGCAGTACTCCAATTTTAGGAG	4967	
TACAGAAACCAACAGACAGTGGAGGTTAGTGCAAGATCTCAGGATTATCAATGAGGCTGTTGTTCTCTAT	5039	domaine
AGCCAGCTGTACCTAGCCCTTACTCTGCTTTCCCAATACAGAGGAAGCAGAGTGGTTTACAGTCTCTGG	5111	pol
ACCTTCAGGATGCCTTCTCTGCATCCCTGTACATCCTGACTCTCAATTCTGTTTGCTTTGAAGTACTT	5183	

FIGURE 1.1

CAAACCCACATCTCAACTCACCTGGACTATTTTACCCCAAGGGTTT	5255	
AGGCATTAGCCCAAGACTTGTAGCCCAATCCTCATACCTGGACACTT	5327	
TTGGCCGCCCATTCAGAAACCTTTGTGCCATCAAGCCACCCAGCGCT	5399	
TACATGGTTTCCAAACCAAGGCTCAACTCTGCTCACAGCAGGTTACT	5471	
CCAGGGCCCTCAGTGAGGAACACATCCAGCCTATACCTGGCTTATC	5543	
GGGGATTCTTTGGCGTAATAGGTTTCTGCCGAAATGGATTCCAGGT	5615	
ATACACTAATTAAGGAACTCAGAAAGCCAAATACCCATTTAGTAAG	5687	
TCCAGGCCCTAACCCAAAGCCCAAGTGTAAAGTTTGGCAACAGGGC	5759	
AAACAGGAATAGCTCTAGGAGTCTTACACAGATCCGAGGGATGAGC	5831	
AGGAAATGTAGTGGCAAAGGGTTGACCTCATTGTTTACGGGTAGTG	5903	
CTGAAGCAGTTAAATAATACAGGGAAGAGATCTTACTGTGTGGACA	5975	
CTGCTAAAGGAGACTTGTGGCTGTGAGCAACTGTTTACTTAAATGT	6047	
TGCTGGCACTGTGCACTTGTGCAACTCTTAAACCCAGCCACATTT	6119	
ATAACTGTCAAGCAAGTAATTTCTCAACCTATGCCATCGAGGGAC	6191	
CCGACCTCAACTTGTATCTGATGGAAGTTCCTTTGTAGAAAAGGAC	6263	
TCAGTGATAATGGAATACTTGAAGTAATCCCTCACTCCAGGAAC	6335	
CTGCACTTGGGCACTAGAAATAGGAGAAGAAAAGGGCAATATATA	6407	
TAGTCTCCATGCCCATGCAGCAATATGGAAGAAAGGGAATTCCTA	6479	
ATCAGGAAGCCATTAGGAAATTTATTTGGCTGTACAGAAACCTAA	6551	
GTCACTCAGAAAGGAAGAAAGGGAATAGAGAACTGCAAGCAGAT	6623	
GGCAGGACCCTCCATTAGAAATGCTTATAAAACAACCCCTAGTAT	6695	
CCCAGTACTCAGCAGGAGAAACAGAAATGGGGAACCTCAGGAGCA	6767	
ACTGAAGAAGGAAATTAATTTTGGCTGCAACTATCCAAAGGAAAT	6839	
CACCTTAGGCATCGATAGCACCATCAGATGGCCAAATCATTTACT	6911	
AAGCAGATAGTCAGGGCTGTGAAGTGTGCCAGAGAAATATCCCT	6983	
GAACAAAGAAGGACCCATTACCTTGGAGAAGACTGGCAACTGAT	7055	
TTCAATCTACTAGTCTGGGTAGATACTTTACCGGGTTGGGCAGAG	7127	
CCCAAGAGGTAATAAGGCACTAGTTCATGAATAATTTCCAGATT	7199	
ACAATAGCCCTGCTTTCCAGGCCACAGTAACCCAGGGAGTATCC	7271	
ACTGCGCTGAAGGCCACAGTCTCAGGGAAGGTCGAGAAATGAAT	7343	
AGCAAAACCCAGGAAACCCACCTCACATGGCCTGCTCTGTTGCC	7415	
CCCAAAAGCAGGACTTAGCCCATACGAAATGCTGTATGGAAGGCC	7487	
ACCCAAGACAGCCAACTTAGTTGCGACATCACCTCTTAGCCAAAT	7559	
AGGAACCTATCCCTGAGAAAGGGGAAAGAACTATTCACCCCTGT	7631	
CTCTAATTCCCCATCCCTAGATACATCTTGGGAAGGACCCTACCC	7703	
TAAAGTGGCTGGAGTGGAGTCTTGGATACATCACTTGAATCCT	7775	
AAATCCAGGAAAGCAAGCTAGCTATTCCTGTGAACCTTAGAGAT	7847	
AGGAGGAAAGTAATAATAATTCATAAATCCCATGGCCCTCCCTT	7919	
ACCCCTCTTCACTCTCACTGCACCCCTCCATGCCCTGTATGACC	7991	
ATGGAGAATGCAAGCTCCCGGAAATATGATGCCCATCGTATAGG	8063	
CACCTGCCACACCCATATGCCCCGCAACTGCTATCACTCTGCCAC	8135	
TTGGACAGGAAATGATTAATCCTAGTGTCTGGAGGACTTGGAGT	8207	
CCAACTGGTATGTCTGATGGGGTGGAGTTCAGATCAGGCAAGGA	8279	
CCAACCTACCCCGGTACATGGCACCTTAGCCCTTACAAAGGACT	8351	domaine
CCGTACCCATACCTCGCCTGGTAAGCTATTTAATACCACCTCACT	8423	
CCCTACTAATCTGTGGATATGCTTCCCTCCCTGAACTCAGGCAT	8495	env
GAACAACCTCAGCAGGAAATAAACACCACTCCGTTTATAGGAC	8567	
CCATACCTCAAACTCACCCTGTGTAAATTTAGCAATACTACATAC	8639	
GGTAACCTCTCCACACAAATAGTCTGCCTACCTCAGGAATATTT	8711	
TTGTTTGAATGGCTCTTTCAGAAATCTATGTGCTTCTCTCACT	8783	
ACAAGATTTATACAGTTATGCTATATCAAGCCCTGCAAAAGAGT	8855	
AGCAGGAGTGTAGGTGCACTAGGTACTGGCATTGGGGTATCACA	8927	
ATCTCAAGAACTAAATGGGGACATGGAACGGGTGCGCGACTCC	8999	
CCTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTTGT	9071	
ATTTTATAGGGGAAGAAATGCTGTTTATGTTAATCAATCCGGA	9143	
AGATCGAATACACGTAGAGCAGAGGAGCTTCGAAACACTGGACC	9215	
CTGGATTCTCCCTTCTTAGGACCTTAGCAGCTATAATTTGCTAC	9287	
CTCTCTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAACT	9359	
GACTAAGATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGAT	9431	
CCCTCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCAAT	9503	
TCGGCCAACCTCCCCAACAGCACTTAGGTTTCTGTTGAGATGGG	9575	
TTCTTAGCTGACTAAGAACTCCCTAAGCCTAGCTGGGAAGGTGAC	9647	
CAACTTAGCTCAGACCTGACCAATCAGAGAGCTCAATAAATGCT	9719	
ATAGCCAATCATCTATGCTGAGAGCAGCAGGAGGACAAATGAT	9791	
CCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGC	9863	region
TAAATCTGCACTGCACTCTTCTGCTCATGTTTCTACGGCTGAG	9935	répétée
ACTGCTGTTTGGCGCCACCGCAGACCCGCGCTGACTCCCTCTG	10007	R1
CTCTGATCCAGCGGACCCATTTGCCGCTCCCAATCGGGCTAAAG	10079	
AGTGCCTGGGTTTCATCTAATGAGCTGAACACTAGTCACTGGGT	10151	
CTTCTAATAGAGCTATAACACTCACCAGATGGCCCAAGGTTCC	10223	
CAGGTACAGAAACAGAGGCTTGCACCATTTGGGAGCTCTGTGAG	10295	
TGAGGTTGCAAAATGCATGGGCCACTAATGGTAGAGCAAGAAAC	10367	
AGTGAGCTGAATGCCTGCCCTGGATGCTTATCTAGGTGTTTCT	10439	
GTTCACTTCTCAAGTAGGGCTTCTATTACAGCCCAATCAATCCC	10500	

FIGURE 1.2

3/15

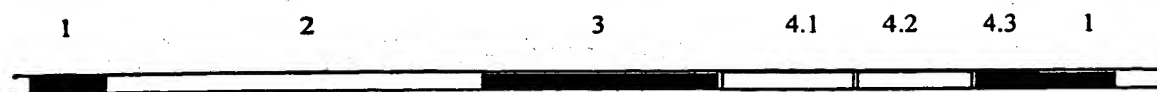


FIGURE 2

ACTGAGAGACAGGACTAGCTGGATTTCCTAGGCCGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC
: : : : :
ACTGAGAGACAGGACTAGCTGGATTTCCTAGGC TGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC
ACGTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGC
: : : : :
ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGC
TAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACA
: : : : :
TAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCTGAGAGCACAGCAGGAGGGACA
ACAATCGGGATATAAACCCAGGCATTGAGCTGGCAACAGCAGCCCCCTTTGGTCCCTTCCCTTTGTA
: : : : :
ATGATCGGGATATAAACCCAAGTCTTCGAGCCGGCAACGGCAACCCCC-TTTGGGTCCCCTCCCTTTGTA
TGGGAGCT--GTTTT CATGCTATTTCACTCTATTAAATCTTGCAACTGCACCTCTCTGGTCCATGTTTCT
: : : : :
TGGGAGCTCTGTTTT CATGCTATTTCACTCTATTAAATCTTGCAACTGCACCTCTCTGGTCCATGTTTCT
TACGGCTCGAGCTGAGCTTTTGCTCACCGTCCACC ACTGCTGTTTGCCACCACCGCAGACCTGCCGCTGA
: : : : :
TACGGCTGAGCTGAGCTTTTCGTTCGCCATCCACC ACTGCTGTTTGCCGCCACCGCAGACCGCCGCTGA
CTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGCGCCCATGCCGCTCC
: : : : :
CTCCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAGGACCCATTGCCGCTCC
CAATTGGGCTAAAGGCTTGCCATTGTTCTCTGCACGGCTAAGTGCTGGGTTTGTTCTAATTGAGCTGAAC
: : : : :
CAATCGGGCTAAAGGCTTGCCATTGTTCTCTGCATGGCTAAGTGCTGGGTTTCATCCTAATTGAGCTGAAC
ACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACA
: : : : :
ACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACAGCTTCTAATAGAGCTATAACACTCACC GCA
TGGCCCAAGATTCCATTCCCTTGGAATCCGTGAGGCCAAGAACTCCAGGTCAGAGAAACGAGGCTTGCCA
: : : : :
TGGCCCAAGGTTCCATTCTTG-AATCCATAAGGCCAAGAACCCAGGTCAGAGAACACGAGGCTTGCCA
CCATCTTGGAAGC
: : : : :
CCATCTTGGGAGC

FIGURE 3

IPMALPYHIFLFTVLLPSFTLTAPPPCRCMTSSSPYQEFLLWRMQRPGNIDAPSYRSLSKG
 TPTFTAHTHMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVWCWYFTQTGMSDGG
 GVQDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTTLTGLHEV
 SAQNPTNCWICLPLNFRPYVSIIPVEQWNNFSTEINTTSVLVGPLVSNLEITHTSNLTCV
 KFSNTTYTTNSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSESMCFLSFLVPPMT
 IYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDM
ERVADSLVTLODQLNSLAAVVLQNRRAIDLLTAERGGTCLFLGEECCYYVNQSGIVTEKVKEIRDRIQRR
AEELRNTGFWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNEVSSRIEAVKLQMEPKMQSKTKIY
RRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSAGSS

FIGURE 4

- 1) NSLAAVVLQNRRAIDLLTAESGGTFLFLEEK
- 2) NSLAAVVLQNRRAIDLLTAERGGTCLFLGEEC
- 3) DSLAAVTLQNHQGLDILLTAEGGLCYFLGEDC
- 4) DSLAAVTLQNHQGLDILLTAEGGLCTFLGEEC
- 5) DSLAAVTLQNCRGLDILLTAEGGHHYTFLEEC
- 6) LQNRRLDILLFLKEGGLC
- 7) DSLAKVVLQNRRLDILLTAEQGGICLALQEK

FIGURE 5

TSFVEKANGVKCHKYKLSFHKETTHNYKSVIYALQEAFRVYLPILPASPTPSPTNKDPPSTQMVQKEIDKRVNS
 EPKSANIPQLXPLQAVGGREFGPARVHVFFSLPDLKQIKTDLGKFSNDNPDGYIDVLQGLGQFFDLTWRDIMSLN
 QTLTPNERSATITAAXEFGDLWYLSQVNDRTTEEREXFPTGQAVPSLDPHWDTESEHGDWCCRHLITCVLEGL
 RKTRKKSMMYSMMSTITQGREENPTAFLERLREALRKRAKSLSPDSSEGQLILKRKFITQSAADIRKKLQKSAVGP
 EQNLETLLNLATSVFYNRDQEEQAEQDKRDXKKGHRFSHDPQASGLWRLWKREKLKGLNAXXGLLPVRSTRTLXK
 RLSKXXKAAPSSMPLISRESLEGPLPQGTKVLXVRSHXPD/SSSRT

FIGURE 6

[illegible]

FIGURE 7

```

01/      TAAATCCCCATGGCCCTCCCTTATCATATTTTCT
02/      TAAATCCCC-TGGCCCTCCCTTATCATATTTTCT
03/      TAAATCCCCATGGCCCTCCCTTATCATATTTTCT
04/      TAGATCCTCATGGCCCTCC-TTGTATATTTTTTT

01/CTTTACTGTTCTTTTA-CCCTCTTCACTCTCACTGCACCCCTCCATGCCGCTGTATGACC
02/CTTTACTGTTCTCTTACCCCCCTTCACTCTCACTGCACCCCTCCATGCCACTGCACCCCTC
03/CTTTACTGTTCTCTTA-CCCCCTTCTCTCACTGCACCCCTCCATGCTGCTGTACAACC
04/CTTTACTGTTCTCTTA-CCCCCTTCACTCTCACTGAACCCCTCCATGCCACTGTACTACC

01/AGT-----AGCTCCCCTTACCAAGAGTTTCTATGGAGAATGCAGCGT
02/GTCCATGCCCGTCTCATGCCAGTAGCTCCCCTTAGCAAGAGTTTCTATGGAGAATGCAGCGT
03/AGC-----AGCTCCCCTTACCAAGAGTTTCTATGAAGAATGCGGCTT
04/AGT-----AGCTCCCATACCAAGAGCTTCTATGGACAATGCGGCTT

01/CCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCTTCACTGC
02/CCCGGAAATATTGATGCCCCATTGTATAGGAGTTTATCTAAGGGAACCCCTTCACTGC
03/CCCAGAAATATTGATGCCCCATCAATAGGAGTTTACCTAAAGGAACTCCACCTTCACTGC
04/CCTGGAAATATTGATGACCCATCGTATAGGAGTTTTTCTAAGGGAACCCCTTCACTGC

01/CCACACCCATATGCCCCGCAACTGCTATCACTCTGCCACTCTTTGCATGCATGCAAATACTC
02/CCACACCCATATGCCCCACAAGTCTATACTCTGCCACTCTTTGCATGCATGCAAATACTC
03/CCACACCCATATGCCCCACAAGTCTATACTCTGCCACTCTTTGCATGCATGCAAATACTC
04/CCACACCTATATGACCC-----

01/ATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGTCACTGTCTGT
02/ATTATTGGACAGGAAAAACGATTAATCCAGTTGTCCTGGAGGACTTGGAG-----
03/ATTATTGGACAGGAAAAATGATTAATCCTAGTTGTCCTGGAGGACTTGGAGCCACTGTCTGT
04/-----

01/TGGACTTACTTCACCCAACTGGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAGA
02/--GACTCACTTCACTCATACCAGTATGTCTGATGGGGGTGGAGTTCAAGATCAGGCAACAGA
03/CGGACTTACTTCACCCATACTGGTATGTCTGAGGGGGTGGAGTTCAAGATCAGGCAAGAGA
04/-----

01/AAAACATGTAAAAGAAGTAATCTCCCAACTCACCCGGGTACATGGCACCTCTAGCCCCCTACA
02/AAAACACATAAAGGAAGTAATCTCCCAACTGACCTGGGTACATAGCACCCCTGGCCCCCTACA
03/AAAACATGTAAAGGAAGTAACCTCCCAACTGACCCGGGTACATAGCACCCCTAGCCCCCTACA
04/-----

01/AAGGACTAGATCTCTCAAACTACATGAAACCCTCCGTACCCATACTCGCCTGGTAAGCCTA
02/AAGGACTAGATCTCTCAAACTACATGAAACCCTCCATACCCATACTGGCCTGGTAAGCCTA
03/AAGGACTAGATCTCTTAAACTACATGAAACCCTCCATACCCATACTTGCCTGGTAAGCCTA
04/-----

01/TTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCCCAAACCCTACTAACTGTTGGAT
02/TTTAATACCACCCTGACTGGGCTCCATGAGGTCTCGGCCCAAACCCTACTAACTGTTGGAT
03/TTTAATACCACCCTCACTGGGCTCCATGAGGTCTCGGTCCAAACCCTACTAACTGTTGGTT
04/-----

01/ATGCCTCCCCCTGAACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACAATGGAACAAC
02/GTGCTCCCCCTGCACTTTAGGCCATACATTTCAATCCCTTATACCTGAACAATGGAACAAC
03/GTGCTCCCCCTGTATTTAGGCCATGCATTTCAATCCCTGTACCTGAACAATGGAACAAC
04/-----TGCACCTCAGGCCATACATTTCAATCCCTGTA-----

```

FIGURE 8.1

01/TCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATA
 02/TCAGCACAGAAATAAACACCACTTCTGTTTTAGTAGGTCCTC---TTTCCAATCTGGAAATA
 03/ACAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTTTCCAATCTGGAAATA

 01/ACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTACATACACAACCAACTCCCA
 02/ACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTATAGACACAGCCAACCTCCCA
 03/ACCCATACCTCAAACCTCACCTGTGTAAAATTTAGCAATACTGTAGACACAACCAACTCCCA
 04/-----
 01/ATGCATCAGGTGGGTAACTCCTCCCACACAAATAGTCTGCCTACCCTCAGGAATATTTTTTG
 02/ATGCATCAGGTGGGTAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTG
 03/ATGCATCAGGTGGGTAACTCCTCCCACACGAATAGTCTGCCTACCCTCAGGAATATTTTTTG
 04/-----
 01/TCTGTGGTACCTCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCA
 02/TCTGTGGTACCTCAGCCTATCATTGTTTGAATGGCTCTTCAGAATCTGTGTGCTTCCTCTCA
 03/TCTGTGGTACCTTAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTGCTTCCTCTCA
 04/-----
 01/TTCTTAGTGCCCCCTATGACCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAA
 02/TTCTTAGTGGCCCCCTATGCCCCATCTACACTGAACAAGATTTATACAATCATGTCATACCTAA
 03/TTCTTAGTGCCCCC-ATGACCATTTACACTGAACAAGATTTATACAATTATGTTGTACCTAA
 04/-----
 01/GCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTATAGGAGCAGGAGTGCTAGGTGCAC
 02/GCCCCGCAACAAAAGAGTACCCATTCTTCCTTTTGTATTGGAGCAGGAGTGCTAGGCGGAG
 03/GCCCCACAACAAAAGAGTACTCATTCTTCCTTTTGTATCGGAGCAGGAGTGCTAGGTGGAC
 04/-----
 01/TAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAA
 02/TAGCTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAAACTGTCTCAAGAA
 03/TAGGTTCTGGCATTGGCGGTACCACAACCTCTACTCAGTTCTACTACAAACTATCTCAAGAA
 04/-----
 01/CTAAATGGGGACATGGAACGGGTGCGCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTC
 02/CTTAAAGGTGACATGGAATGGGTGCGTGATACCCTGGTCACCTTGCAAGATCAACTTAACTC
 03/CTCAATGGTGACATGGAATGGGTGCGCGACTCCCTGGTCACCTTGCAAGATCAACTTAACTT
 04/-----
 01/CCTAGCAGCAGTAGTCCTTCAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGG
 02/CCTAGCAGCAGTAGTCCTTCAAATCGAAGAGCTTTAGACTTGCTAACCGCGGAAAGCGGGG
 03/CCTAGCATCAGTAGTCCTTCAAATCGAAGAGCTTTAGACTTGCTAACCTCTGAAAGAGGGG
 04/-----
 01/GAACCTGTTTATTTTTAGGGGAAGAATGCTGTTATTATGTT-----
 02/GAACCTTTTTATTTTTAGAGGAAAAATGCTGTTGTTATGTT-----
 03/GAAGCTGTTTATTTTTAGGGGAAGAATGTTGTTATTATGTTATTTTAGCGGAAGAATGTTGT
 04/-----
 01/-----AATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACA
 02/-----AATCAATCCGGAATCATACCGAGAAAGTTAAAGAAATTCAGGTCGAATATA
 03/TATTATGTTAATCAATCCTGAATTGTCACAGAGAAAGTTGAAGAAATTCGAGATTGAATACA
 04/-----
 01/ACGTAGAGCAGAGGAGCTTCGAAA-CACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCT
 02/ACGTAGAGCAAAGGAGCTGCAAAA-CACTGGACCCTGGGGCCTCCTCAGCCAATGGATGCCCT
 03/ACGTAGAACAGAGGAGCTTCAAAAACACCAGACCCTGGGGCCTCCTCAGCCAATGGATGCCCT
 04/-----

FIGURE 8.2

9/15

01/GGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTA
02/GGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGTTACTCCTCTTTGGACCCTGTA
03/GGATTCTCCCCTTCTTAGGATCTCTAGCAGCTCTAATATTGATACTCCTCTTTGGACCCTGTA
04/-----

01/TCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTAAACTA-----
02/TCTTTAACCTCCTTGTTAAGTTTGTCTTTTCCAGAATCGAAGCAGTAAACTACAAATCGTTC
03/TCTTTAACCTCCTTGTTAAGTTTGTCTCTTCCAGAATCAAAGTTGTAAAGCTACAAATCGTTC
04/TCTTTAACCTCCTTGTTAAGCTTGTCTCTTGCAGAATCGAAGCTGTAAACTACAAATGCTTG

01/--CAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCCTG
02/TTCAAATGGAGCCCCAGATGCAGTCCATGAGTAAATCTACCACGGACCCCTGGACCGGCCTG
03/TTCAAATGGAACCCCAGATGAAGTCCATGACTAAGATCTACCGTGGACCCCTGGACCGGCCTA
04/TTAAATAGAGCCCCAGATGCAGTCCATGGCTAAGATCTACCACGGACCCCTGGACCGGCCTG

01/CTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAGCTGCAC
02/CTAGCCCATGCTCTGATGTTAATGACATCAAAGGCACCCCTCCCGAGGAAATCTCAACTGCAC
03/CTAGCCCATGCTCCAATTGTAATGATATCGAACGCACCCCTCCCGAGGAAATCTCAACTGCAC
04/CTAGCCCATGCTCTGATGTTGATGACATTGAAGGCACGGCTTCGAGGAAATCTCAACTGCAC

01/AACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCGTCGGCCAACCTCCCC
02/AACCTCTACTACGCCCCAATTCAGCAGGAAGCAGTTAGAGTGGTTGTTGGCCAACCTCCCC
03/AACCCCTACTATGCCCCAATTCGCGAGGAAGCAGTTAGACTGGTCGTCAGCCAACCTCCCC

04/GACCCCTACTACACCCCAATTTAGCGGGAAGCAATTAGAGCAGCCTATGGCCACCTCCCC

FIGURE 8.3

CTTCCCCAACTAATAAGGACCCCCCTTTCAACCCAAACAGTCCAAAAGGACATAGACAAAGGA	3
CTTCCCCAACTAATAAGGACCCCCCTTTCAACCCAAACAGTCCAAAAGGACATAGACAAAGGA	4
CTTCCCCAACTAATAAGGACCCCCCTTTCAACCCAAATGGTCCAAAAGGAGATAGACAAAGG	5
CTTCTCCAATAATAAGGACCCCCCTTTCAACCCAAATGGTCCAAAAGGAGATAGACAAAGG	6
CTTCCCCAAATAATAAGAACCCCCCTTTCAACCCAAACGGTCCAAAAGGAGATAGACAAAGG	7
GTAAACAATGAACCAAAGAGTGCCAATATTCCCTGGTTATGCACCCTCCAAGCGGTGGGAG--	3
GTAAACAATGAACCAAAGAGTGCCAATATTCCCTGGTTATGCACCCTCCAAGCGGTGGGAG--	4
GTAAACAGTGAACCAAAGAGTGCCAATATTCCCCAATTATGACCCCTCCAAGCAGTGGGAGGA	5
GTAAACAATGAACCAAAGAGTGCCAATATTACACGATTATACTCGCTCCAAGCAGTGGGAG--	6
GTAAACAATAACCAAAGATGCCAATATTCCCCGATTATGCCCCCTCCAAGCGGTGGGAG--	7
A-AGAATTTCGGCCCAGCCAGAGTGCATGTACCTTTTTCTCTCTCAC-ACTTGAAGCAAATTAAA	3
A-AGAATTTCGGCCCAGCCAGAGTGCATGTACCTTTTTCTCTCTCAC-ACTTGAAGCAAATTAAA	4
AGAGAATTTCGGCCCAGCCAGAGTGCATGTGCCTTTTTCTCTCCCAG-ACTTAAAGCAAATAAAA	5
-GAGAATTTGGCCCAGCCAGCGTGCATGTACCTTTTTCTCTCTCAG-ATTTAAAGCAAATTAAA	6
-GAGAATTTCGGCCCAGCCAGAGTGCACGTACCTTTTTCTCTCTCTAGACTTTAAA-----TTAAA	7
ATAGACNTAGGTNAATTNTCAGATAGCCCTGATGGYTATATTGATGTTTTACAAGGATTAGGA	3
ATAGACXTAGGTXAATTXTCAGATAGCCCTGATGGXTATATTGATGTTTTACAAGGATTAGGA	4
ACAGACTTAGGTAAATTCTCAGATAACCCTGATGGCTATATTGATGTTTTACAAGGGTTAGGA	5
ATAGACCTAGGTAAATTCTCAGATAACCCTGATGGCTATATTGATGTTTTACAAGGGTTAGGA	6
ATAGACCTAGGTAAATTCTCAGATAACCCTAATGGCTATATTGATGTTTTACAAGGGTTAGGA	7
TTCTTGAGTTCTTGCACTAACCTCAAAT	1
CAATCCTTTGATCTGACATGGAGAGATATAATATTACTGCTAAATCAGACGCTAACCTCAAAT	3
CAATCCTTTGATCTGACATGGAGAGATATAATATTACTGCTAAATCAGACGCTAACCTCAAAT	4
CAATTCCTTTGATCTGACATGGAGAGATATAATGTCACTGCTAAATCAGACACTAACCCCCAAAT	5
CAATCCTTTGATCTGACATGGAGAGATATAATGTTACTGCTAAATCAGACACTAACCCCCAAAT	6
CAATCCTTTGATCTGATATGGAGAGATATAATGTTACTGCTAAATCAGACACTAACCCCCAAAT	7
GAGAGAAGTGCCGCATAAAGTGAACCCAAAGAGTTTGCGCATCCCTGGTATCTCAGTCAGGTC	1
GAGAGAAGTGCTGCCATAAAGTGGAGCCCGAGAGTTTGGAATCTCTGGTATCTCAGTCAGGTC	3
GAGAGAAGTGCTGCCATAAAGTGGAGCCCGAGAGTTTGGAATCTCTGGTATCTCAGTCAGGTC	4
GAGAGAAGTGCCACCATAAAGTGCAGCCTGAGAGTTTGCGCATCTCTGGTATCTCAGTCAGGTC	5
GAAAAAAGTGCTGCCATAACAGCAGCCTGAGAGTTTGCGGAATCTCTGGTATCTCAGTCAGGTC	6
GACAGAAGTGTCGCGGTAAGTGGAGCCCGAGAGTTTGGAATCTCTGGTATCTCAGTCAGGTC	7
AATGACAGGATGACAACAGAGGAAAGATAATGATTCCCCACAGGCCAGCAGGCAGTTCCCAGT	1
AATGATAGGATGACAACGGAGGAAAGAGAACGATTCCCCACAGGGCAGCAGGCAGTTCCCAGT	3
AATGATAGGATGACAACGGAGGAAAGAGAACGATTCCCCACAGGGCAGCAGGCAGTTCCCAGT	4
AATGATAGGATGACAACAGAGGAAAGAGAATGATTCCCCACAGGCCAGCAGGCAGTTCCCAGT	5
AATGATAGGATGACAACAGATGAAAGAGAATGATTCCCCACAGGCCAGCAGGCAGTTCCCAGT	6
AATGATAGGATGACAACAGAGGAAAGAGAACGATTCCCCACAGGCCAGCAGGCAGTTCCCAGT	7
GTAGACCCTCATTAGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTTGCTAACT	1
AACT	2
GTAGCTCCTCATTGGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTTACTAACT	3
GTAGCTCCTCATTGGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTT	4
CTAGACCCTCATTGGGACACAGAATCAGAACATGGAGATTGGTGCTGCAGACATTTGCTAACT	5
GTAGACCCTCATTAGGACACAGAATCAGAACTTGGAGATTGGTGCCACAGACATTTGCTAACT	6
GTAGACCCTCACTGGGACACAGAATCAGAACATGGAGATTGGTGCCGCAGACATTTGCTAACT	7

FIGURE 9.1

TGCGTGCTAGAAGGACTAAGGAAAAC TAGGAAGA----	1
TGCGTGCTAGAAGGACTAAGGAAAAC TAGGAAGA---CTATGAATTATTCAATGATGTCCACT	2
TGCGTGCTAGAAGGACTAAGGAAAAC TAGGAAGA---CTATGAATTATTCAATGATGTCCACT	3
TGTGTGCTAGAAGGACTAAGGAAAAC TAGGAAGAAGTCTATGAATTACTCAATGATGTCCACA	5
TGCGTGCTAGAAGGACTAAGGAAAAC TAGGAAGAAGCCCATGAATTATTCAATGATGTCCCT	6
TGCGTGCTAGAAGGACTAAGGAAAAC TAGAAAGAAGCCTGTGAGTTATTCAATGATGTCCACT	7
ATAACACAGGGGAAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	1
ATAACACAGGGGAAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	2
ATAACACAGGGGAAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	3
ATAACACAGGG-AAGGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAG	5
ATAACACAGGG-AAAGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAAGGATTGAG	6
ATAACACAGGG-AAAGGAAGAAAATCCTACCGCCTTTCTGGAGTGACTAACGGAGGCATTGAG	7
GAAGCATACC---AGGCAAGTGGACATTGGAGGCTCTGGAAAAGGGAAAAGTTGGGAAAAGTA	1
GAAGCATACC---AGGCAAGTGGACATTGGAGGCTCTGGAAAAGGGAAAAGTTGGGCAAATTG	2
GAAGCATACC---AGGCAAGTGGACATTGGAGGCTCTGGAAAAGGGAAAAGTTGGGCAAATTG	3
GAAGCGTGCC232AGGCAAGTGGACTTTGGAGGCTCTGGAAAAGGGAAAAGCTGGGCAAATTG	5
GAAGCATACC238AGGCAAATGGACTTTGGAGGCTCCAGAAAAGGGAAAAGCTGAGCAAATTG	6
GAAGCATACC233AGGCAAGCGGACTTTGGAGGCTGGA AAAAGGGAAAAGCTAGGCAAATCA	7
TATGTCTAATAGGGCTTGCTTCCAGTGTGGTCTACAAGGACACTTTAAAAAAGATTGTCC-AA	1
AATGCCTAATAGGGCTTGCTTCCAGTGCAGTCTACAAGGACGCTTTAGAAAAGATTGTCC-AA	2
AATGCCTAA	3
AATGCCTAATAGGGCTTGCTTCCAGTGCAGTCTACAAGGACACTTTAAAAAAGATTGTCC-AA	5
AATGCCTAACAGGGCTTGCTTCTAGTGTGGTCTACAAGGACACTTTAAAAAAGATTGTCC-AA	6
AATGCCTAATAGGGTTTGCTTCCAGTGCAGTCTACAAGGACACTTTAAAAAAGATTGTCCAA	7
-TAGAAATAAGCCACCACCTCGTCCATGCCCTTATGTCAAGGGAATCACTGGAAGGCCCACT	1
G TAGAAATAAGCCGCCCC-TCGTCCATGCCCTTATGTCAAGGGAATCACTGGAAGGCCCTACT	2
G TAGAAGTAAGCCGCCCCCTCGTCCATGCCCTTATTTCAAGGGAATCACTGGAAGGCCCACT	5
G TAGAAACAAGCTGCCCCCTTGTCATGCCCTTATGTCAAGGGAATCACTGGAAGGCCCACT	6
-TAGAAATAAGCCGCCCCCTCGTCCATGCACCTCGTGTCAAGGGAATCACTGTAAGGCCCACT	7
GCCCCAGGGGATGAAGGTCTCTGAGTCAGAAGCCACTAACCAGATGA	1
GCCCCAGGGGACGAAGGTCTCTGAGTCAGAAGCCACTAACCTGATGA	2
GCCCCAGGGGACAAAGGTCTCTGAGTCAGAAGCCACTAACCAGATGA	5
GCCCCAGGAGATGAAGGTCTCTGAGTCAGAAGCCACTAACCAGATAA	6
GCCCCAGGGGACGTAGGTCTCTGAGTCAGAAGCCACTAACCAGATGA	7

FIGURE 9.2

12/15

RTPLSTQTVQKDIDKGVNNEPKSANIPWLCTLQAVGEEFGPARVHVPFSLSHLKQIKIDG SDSPDG
- = == ----- == ===== = ===== ===== = = == ==
KDPSTQMVQKEIDKRVNSEPKSANIPQLPLQAVGGREFGPARVHVPFSLPDLKQIKTDLGKFSDNPDG

YIDVLQGLGQSFDLTWRDIILLNQTLSNERSAAITGAREFGNLWYLSQVNDRTTEERERFPTGQQ
===== ----- ===== ----- ===== =====
YIDVLQGLGQFFDLTWRDIMSLLNQTLPNERSATITAA~~X~~EFGDLWYLSQVNDRTTEEREX~~F~~PTGQQ

AVPSVAPHWDTESEHGDWCCRHLCTVLEGLRKTRK TMNYSMMSTITQ GK

AVPSLDPHWDTESEHGDWCCRHLCTVLEGLRKTRKSMNYSMMSTITQGR

FIGURE 10

TTCCTGTTGAGATGGGG

FIGURE 11

[illegible]

FIGURE 13

